STIC-Biotech/ChemLib 116874

From: Chan, Christina
Sent: Monday, March 15, 2004 8:17 AM
To: Ford, Vanessa; STIC-Biotech/ChemLib
Subject: RE: In re: 09/543407 sequence search

Importance: High

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----
Fr m: Ford, Vanessa
Sent: Friday, March 12, 2004 1:50 PM
To: Chan, Christina
Subject: In re: 09/543407 sequence search

Please search: SEQ ID NOs: 1, 2, 3, 4, 9, 11, 13, 15, 17, 19. please include inteference searches. Please Rush.

Biotechnology Patent Examiner
Office: REM 3B25
Mailbox: REM 3C18
Phone: 571.272.0857
Art Unit: 1645

Searcher:_____
Phone:_____
Location:_____
Date Picked Up: 3/15/04
Date Completed:_____
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences: 10
AA Sequences:_____
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.: CPST
WWW/Internet:_____
Other (specify):_____

THIS PAGE BLANK (USPTO)

116817

STIC-Biotech/ChemLib

From: Page, Thurman
Sent: Friday, March 12, 2004 2:37 PM
To: STIC-Biotech/ChemLib
Cc: Ford, Vanessa
Subject: FW:

Importance: High

CRF E

Rush Search Approved.

Vanessa,
Please resend the reason for rush.

Thurman K. Page
SPE Art Units 1615 & 1616
Technology Center 1600

-----Original Message-----
From: Ford, Vanessa
Sent: Friday, March 12, 2004 1:55 PM
T : Page, Thurman
Subject: FW:

-----Original Message-----
Fr m: Ford, Vanessa
Sent: Friday, March 12, 2004 1:54 PM
T : Page, Thurman
Subject:

Christina is out of the Office today.

Please search: SEQ ID NOs: 21, 23 , 25, 27 and 29 please include inteference searches. Please Rush. Sequence search Application No. 09/543, 407

Biotechnology Patent Examiner
Office: REM 3B25
Mailbox: REM 3C18
Phone: 571.272.0857
Art Unit: 1645

Searcher:_____
Phone:_____
Location:_____
Date Picked Up:_____
Date Completed:_____
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences:_____
AA Sequences:_____
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.:_____
WWW/Internet:_____
Other (specify):_____

THIS PAGE BLANK (USPTO)

18522

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 19:02:18 ; Search time 2269.8 Seconds
(without alignments)
8707.562 Million cell updates/sec

Title: US-09-543-407-21
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		401.6	88.1	456	6	I44909	I44909 Sequence 58
2		401.6	88.1	2067	1	SEU43280	U43280 Salmonella
3		395.2	86.7	5103	1	STAJ2301	AJ002301 Salmonell
4		395.2	86.7	22411	1	AE008749	AE008749 Salmonell
5		393.6	86.3	254050	1	AL627269	AL627269 Salmonell
6	c	393.6	86.3	301983	1	AE016840	AE016840 Salmonell
7		387.8	85.0	1048	1	STAGFBA	AJ000514 Salmonell
8		277.4	60.8	361	6	I44908	I44908 Sequence 56
9		267.2	58.6	2889	1	CSP515700	AJ515700 Citrobact
10		232	50.9	4680	1	ECCSGABDG	X90754 E.coli csgG
11		232	50.9	10346	1	AE000205	AE000205 Escherich
12		232	50.9	15047	1	D90741	D90741 Escherichia
13		230.8	50.6	2920	1	CFR515701	AJ515701 Citrobact
14		230.4	50.5	456	6	AX814811	AX814811 Sequence
15		230.4	50.5	648	1	ECOCSGAA	L04979 Escherichia
16		222.2	48.7	1711	1	AF275733	AF275733 Escherich
17		222.2	48.7	10190	1	AE005315	AE005315 Escherich
18		222.2	48.7	327773	1	AP002554	AP002554 Escherich
19		217.4	47.7	306358	1	AE016759	AE016759 Escherich
20		184	40.4	2883	1	ESA515702	AJ515702 Enterobac
21		174.6	38.3	230	1	SEU53207	U53207 Salmonella
22		164.8	36.1	10370	1	AE015131	AE015131 Shigella
23		164.8	36.1	292504	1	AE016981	AE016981 Shigella
24	c	141.8	31.1	437	1	AF237726	AF237726 Shigella
25		79.6	17.5	19201	1	D90742	D90742 Escherichia
26		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27	c	49.4	10.8	2000	6	AX655393	AX655393 Sequence
28		48.2	10.6	78	6	AX814809	AX814809 Sequence
29		44.8	9.8	72	6	AX814798	AX814798 Sequence
30		36.8	8.1	926	1	AF193498	AF193498 Vibrio ha
31		36.4	8.0	336	1	SSU491341	AJ491341 Streptoco
32		36.4	8.0	336	1	SSU491343	AJ491343 Streptoco
33		36.4	8.0	336	1	SSU491344	AJ491344 Streptoco
34		36.4	8.0	336	1	SSU491346	AJ491346 Streptoco
35		36.4	8.0	336	1	SSU491347	AJ491347 Streptoco
36		36.4	8.0	336	1	SSU491348	AJ491348 Streptoco
37		36.4	8.0	336	1	SSU491349	AJ491349 Streptoco
38		36.4	8.0	336	1	SSU491350	AJ491350 Streptoco
39		36.4	8.0	336	1	SSU491351	AJ491351 Streptoco
40		36.4	8.0	336	1	SSU491352	AJ491352 Streptoco
41		36.4	8.0	336	1	SSU491357	AJ491357 Streptoco
42		36.4	8.0	336	1	SSU491358	AJ491358 Streptoco
43		36.4	8.0	336	1	SSU491359	AJ491359 Streptoco
44		36.4	8.0	336	1	SSU491367	AJ491367 Streptoco
45		36.4	8.0	336	1	SSU491370	AJ491370 Streptoco

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match              88.1%;  Score 401.6;  DB 6;  Length 456;
  Best Local Similarity    92.5%;  Pred. No. 7.8e-102;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||||| | | | | | | | | | | | | | | | | | | | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match              88.1%;  Score 401.6;  DB 1;  Length 2067;
  Best Local Similarity    92.5%;  Pred. No. 8.3e-102;
```

```
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |      || |    || |  |    | |
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
-10_signal           complement(2869. .2874)
                     /gene="csgDEFG"
-35_signal           complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
CDS             /gene="csgA"
                3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
CDS             /gene="csgC"
                4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
CDS             /gene="ORF103"
                4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN
```

```
  Query Match             86.7%;  Score 395.2;  DB 1;  Length 5103;
  Best Local Similarity   91.7%;  Pred. No. 5.5e-100;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||||  | | |  |  |||  ||  |  |  |  |  |  |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421
```

```
RESULT 4
AE008749
LOCUS       AE008749                 22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT .    Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES            Location/Qualifiers
     source         1. .22411
                    /organism="Salmonella typhimurium LT2"
                    /mol_type="genomic DNA"
                    /strain="LT2; SGSC 1412; ATCC 700720"
                    /db_xref="ATCC:700720"
                    /db_xref="taxon:99287"
                    /note="LT2"
     gene           434. .1308
                    /gene="phoH"
                    /note="synonym: STM1126"
     RBS            434. .439
                    /gene="phoH"
                    /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS            454. .1308
                    /gene="phoH"
                    /note="similar to E. coli PhoB-dependent, ATP-binding pho
                    regulon component; may be helicase; induced by P
                    starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                    aa), 92% identity in aa 71 - 354"
                    /codon_start=1
                    /transl_table=11
                    /product="PhoB-dependent ATP-binding pho regulon
                    component"
                    /protein_id="AAL20057.1"
                    /db_xref="GI:16419642"
                    /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                    RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                    GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                    YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                    LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                    QRTLHAYS"
     gene           complement(1414. .2303)
                    /gene="STM1127"
     CDS            complement(1414. .2295)
                    /gene="STM1127"
                    /note="similar to E. coli orf, hypothetical protein
                    (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                    identity in aa 15 - 264"
                    /codon_start=1
                    /transl_table=11
                    /product="putative transcriptional regulator"
                    /protein_id="AAL20058.1"
                    /db_xref="GI:16419643"
                    /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                    SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                    SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                    MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                    TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                    QARESKLRTMNALDMTK"
     RBS            complement(2298. .2303)
                    /gene="STM1127"
                    /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene           complement(2580. .4088)
                    /gene="STM1128"
     CDS            complement(2580. .4076)
                    /gene="STM1128"
                    /note="similar to E. coli putative cotransporter
                    (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                    identity in aa 7 - 478"
                    /codon_start=1
                    /transl_table=11
                    /product="putative sodium/glucose cotransporter"
                    /protein_id="AAL20059.1"
                    /db_xref="GI:16419644"
                    /translation="MITHSPGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                    VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPPFRK
                    LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                    LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                    DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                    NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                    AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                    VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                    RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS            complement(4083. .4088)
                    /gene="STM1128"
                    /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene           complement(4413. .5099)
                    /gene="STM1129"
     CDS            complement(4413. .5093)
                    /gene="STM1129"
                    /note="similar to E. coli putative enzyme (AAC76255.1);
                    Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                    226"
                    /codon_start=1
                    /transl_table=11
                    /product="putative inner membrane protein"
                    /protein_id="AAL20060.1"
                    /db_xref="GI:16419645"
                    /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                    AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                    AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGVTGP
                    ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                    WFSHAVKR"
     RBS            complement(5094. .5099)
                    /gene="STM1129"
                    /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene           5599. .6759
                    /gene="STM1130"
     CDS            5599. .6759
                    /gene="STM1130"
                    /note="similar to E. coli orf, hypothetical protein
                    (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                    identity in aa 40 - 403"
                    /codon_start=1
                    /transl_table=11
                    /product="putative inner membrane protein"
                    /protein_id="AAL20061.1"
                    /db_xref="GI:16419646"
                    /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                    AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                    GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                    EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                    PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                    QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                    WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene           6793. .7497
                    /gene="STM1131"
     RBS            6793. .6798
                    /gene="STM1131"
                    /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNPDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match          86.7%;  Score 395.2;  DB 1;  Length 22411;
  Best Local Similarity  91.7%;  Pred. No. 5.8e-100;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||||  | |  |||| |    || | |  |  | ||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
  PUBMED    11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAARERIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             86.3%;  Score 393.6;  DB 1;  Length 254050;
  Best Local Similarity   91.4%;  Pred. No. 1.8e-99;
  Matches  417;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||| | | | || | || | | | | |
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||  ||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361
```

RESULT 6
AE016840/c
LOCUS AE016840 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of 16 of the complete genome.
ACCESSION AE016840 AE014613
VERSION AE016840.1 GI:29137797
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi Ty2
ORGANISM Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 301983)
AUTHORS Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and CT18
JOURNAL J. Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE 22531367
PUBMED 12644504
REFERENCE 2 (bases 1 to 301983)
AUTHORS Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES Location/Qualifiers

```
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
     CDS             /locus_tag="t1735"
                     complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
     CDS             /locus_tag="t1736"
                     complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
     gene            NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
                     complement(1812. .3473)
                     /gene="flgK"
     CDS             /locus_tag="t1737"
                     complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
     gene            QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
                     complement(3538. .4488)
                     /gene="flgJ"
     CDS             /locus_tag="t1738"
                     complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
     gene            NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
                     complement(4488. .5591)
                     /gene="flgI"
     CDS             /locus_tag="t1739"
                     complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEPIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRPFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.3%;  Score 393.6;  DB 1;  Length 301983;
Best Local Similarity   91.4%;  Pred. No. 1.8e-99;
  Matches  417;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||||  | |  |  || |  || | |  | |
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||     ||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
     gene            KTAVVVQKQSHMAIRVTQR"
                     593. .1048
     CDS             /gene="agfA"
                     593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             85.0%;  Score 387.8;  DB 1;  Length 1048;
  Best Local Similarity   90.8%;  Pred. No. 6.1e-98;
  Matches  413;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||    |  |      | || |   || |  |    | |
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             60.8%;  Score 277.4;  DB 6;  Length 361;
  Best Local Similarity   89.3%;  Pred. No. 6.5e-67;
  Matches  299;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 303
              ||||||||||||||||||  |  |      | || |   || |  |    | | ||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 CATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           58.6%;  Score 267.2;  DB 1;  Length 2889;
  Best Local Similarity  75.7%;  Pred. No. 5.3e-64;
  Matches  345;  Conservative    0;  Mismatches  108;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |     ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||   | |||||| ||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||   ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              || |||||||||  | |||||                | || |     |    | |  | |
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||        || |||||||||  ||||| || || ||||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || ||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
```

```
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
     CDS             /gene="csgE"
                     complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
     gene            IQTEEALNRRQINQALLSTGDLAHDEF"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
     misc_feature    KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
                     complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
     CDS             /gene="csgB"
                     3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
     gene            SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     CDS             3729. .4184
                     /gene="csgA"
                     3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
     sig_peptide     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     gene            3729. .3788
     CDS             /gene="csgA"
                     4243. .4575
                     /gene="orfC"
                     4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     ORIGIN
```

```
  Query Match           50.9%;  Score 232;  DB 1;  Length 4680;
  Best Local Similarity 69.3%;  Pred. No. 4.2e-54;
  Matches  316;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||    | ||| |||||| ||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||||| |||||  ||| |||||||||  ||     ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              || || |||||  | ||| |                | ||||    ||    | | | |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||   |  || ||||||||||  ||||| || || || || ||     || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  ||| |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184
```

```
RESULT 11
AE000205
LOCUS       AE000205                10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
```

```
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
```

```
     promoter        /note="factor Sigma70; predicted +1 start at 1096980"
                     2338. .2365
     promoter        /note="factor Sigma70; predicted +1 start at 1097049"
                     2355. .2383
     gene            /note="factor Sigma70; predicted +1 start at 1097067"
                     2393. .3370
                     /gene="ycdW"
     CDS             /note="synonym: b1033"
                     2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
     promoter        /note="factor Sigma70; predicted +1 start at 1097992"
                     3383. .3410
     promoter        /note="factor Sigma70; predicted +1 start at 1098094"
                     3396. .3423
```

Query Match 50.9%; Score 232; DB 1; Length 10346;
Best Local Similarity 69.3%; Pred. No. 4.4e-54;
Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy       1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db    8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy      61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db    9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy     121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db    9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy     181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db    9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy     241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
Db    9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy     301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db    9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy     361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db    9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy     421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db    9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

```
RESULT 12
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                   Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES                   Location/Qualifiers
     source                1. .15047
                           /organism="Escherichia coli K12"
                           /mol_type="genomic DNA"
                           /strain="K12"
                           /db_xref="taxon:83333"
                           /map="23.7-24.0 min"
                           /clone="Kohara clone #231"
                           /note="Nucleotide position 1097101-1112147 from the
```

```
                initiation site of ThrA (0 min.).~This clone is from
                Kohara lambda miniset library."
     CDS        <1. .276
                /note="unnamed protein product; ORF_ID:o230#5"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35822.1"
                /db_xref="GI:4062605"
                /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
                WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     CDS        1152. .2090
                /note="ORF_ID:o231#1
                similar to SwissProt Accession Number P45637"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical 33.0 kd protein in proB-proA
                intergenic region."
                /protein_id="BAA35823.1"
                /db_xref="GI:4062606"
                /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
                WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
                AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
                PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
                GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
                AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     CDS        2145. .2882
                /note="unnamed protein product; ORF_ID:o231#2"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35824.1"
                /db_xref="GI:4062607"
                /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
                ILNVSPRRLLNFLESRGMAPIAEFADL"
     CDS        2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
     CDS        3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
     gene       complement(4117. .4950)
                /gene="csgG"
     CDS        complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
     gene       complement(4977. .5393)
                /gene="csgF"
     CDS        complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene       complement(5418. .5807)
                /gene="csgE"
     CDS        complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
     gene       complement(5812. .6462)
                /gene="csgD"
     CDS        complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene       7217. .7672
                /gene="csgB"
     CDS        7217. .7672

  Query Match           50.9%;  Score 232;  DB 1;  Length 15047;
  Best Local Similarity  69.3%;  Pred. No. 4.5e-54;
  Matches  316;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952
```

```
Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168
```

RESULT 13
CFR515701
LOCUS CFR515701 2920 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION AJ515701
VERSION AJ515701.1 GI:31790495
KEYWORDS csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation component of curlin monomers; regulatory protein.
SOURCE Citrobacter freundii
ORGANISM Citrobacter freundii
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.
REFERENCE 1
AUTHORS Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
JOURNAL Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE 2 (bases 1 to 2920)
AUTHORS Romling,U.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES Location/Qualifiers
source 1. .2920
/organism="Citrobacter freundii"
/mol_type="genomic DNA"
/isolate="Fec4"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:546"
/country="Germany"
gene complement(211. .861)
/gene="csgD"
CDS complement(211. .861)
/gene="csgD"
/function="necessary for transcription of the csgBA operon"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56673.1"
/db_xref="GI:31790496"
/translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene 1624. .2082
CDS 1624. .2082
/gene="csgB"
/function="nucleator"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56674.1"
/db_xref="GI:31790497"
/translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
gene 2123. .2572
/gene="csgA"
CDS 2123. .2572
/gene="csgA"
/function="curli subunit"
/codon_start=1
/transl_table=11
/product="curlin-csgA protein"
/protein_id="CAD56675.1"
/db_xref="GI:31790498"
/translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN

```
Query Match           50.6%;  Score 230.8;  DB 1;  Length 2920;
Best Local Similarity 70.8%;  Pred. No. 9e-54;
Matches  323;  Conservative    0;  Mismatches  127;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

RESULT 14
AX814811
LOCUS AX814811 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
TITLE Peptides
JOURNAL Patent: WO 03064446-A 15 07-AUG-2003;
Hansa Medical Research Aktiebolag (SE)
FEATURES Location/Qualifiers
source 1. .456

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           50.5%;  Score 230.4;  DB 6;  Length 456;
  Best Local Similarity  69.1%;  Pred. No. 1.1e-53;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              || || |||||  | ||| |             | ||||    ||    |  | | |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||   |   || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | ||| ||| | ||| |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match           50.5%;  Score 230.4;  DB 1;  Length 648;
  Best Local Similarity  69.1%;  Pred. No. 1.1e-53;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              || || |||||  | ||| |               | ||||    ||    |  | | |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||   |   || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | ||| ||| | ||| |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 15, 2004, 07:24:04
Job time : 2276.2 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 18:52:53 ; Search time 290.4 Seconds
(without alignments)
6670.724 Million cell updates/sec

Title: US-09-543-407-21
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64627	Aac64627 AgfA::PT3
	2	401.6	88.1	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	401.6	88.1	456	2	AAT74142	Aat74142 Salmonell
	4	401.6	88.1	456	3	AAC64617	Aac64617 Salmonell
	5	376	82.5	456	3	AAC64630	Aac64630 AgfA::PT3
	6	358.4	78.6	456	3	AAC64626	Aac64626 AgfA::PT3
	7	352	77.2	456	3	AAC64625	Aac64625 AgfA::PT3
	8	350.4	76.8	456	3	AAC64628	Aac64628 AgfA::PT3
	9	350.4	76.8	456	3	AAC64622	Aac64622 AgfA::PT3
	10	350.4	76.8	456	3	AAC64629	Aac64629 AgfA::PT3
	11	350.4	76.8	456	3	AAC64623	Aac64623 AgfA::PT3
	12	348.8	76.5	456	3	AAC64631	Aac64631 AgfA::PT3
	13	348.8	76.5	456	3	AAC64624	Aac64624 AgfA::PT3
	14	277.4	60.8	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	277.4	60.8	361	2	AAT74141	Aat74141 Salmonell
	16	230.4	50.5	456	3	AAC64619	Aac64619 Escherich
	17	230.4	50.5	456	9	ACF36153	Acf36153 E. coli C
	18	218	47.8	646	2	AAQ62647	Aaq62647 Fibronect
	19	156	34.2	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
	21	50.8	11.1	78	3	AAC64610	Aac64610 AgfA (SEF
c	22	50.6	11.1	78	3	AAC64609	Aac64609 AgfA (SEF
c	23	49.4	10.8	2000	7	ADA71938	Ada71938 Rice gene
	24	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	25	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	26	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
	27	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	28	41	9.0	78	3	AAC64606	Aac64606 SefA (SEF
c	29	40.8	8.9	78	3	AAC64605	Aac64605 SefA (SEF
c	30	35.4	7.8	494	4	ABA58813	Aba58813 Human foe
c	31	35.4	7.8	494	4	AAI38517	Aai38517 Probe #72
c	32	35.4	7.8	494	4	AAK32702	Aak32702 Human bon
c	33	35.4	7.8	494	4	AAK06966	Aak06966 Human bra
c	34	35.4	7.8	494	4	ABS32421	Abs32421 Human liv
c	35	35.4	7.8	494	6	ABS07498	Abs07498 Human gen
	36	35	7.7	100	7	ACD68808	Acd68808 E. coli K
c	37	35	7.7	1023	4	ABL11185	Abl11185 Drosophil
	38	35	7.7	2553	3	AAA09500	Aaa09500 V. furnis
c	39	35	7.7	4147	4	ABL11184	Abl11184 Drosophil
c	40	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	41	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	42	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	43	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	44	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	45	34	7.5	1344	6	AAD27016	Aad27016 E. amylov

ALIGNMENTS

RESULT 1
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36351.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

```
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 4e-136;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag=  a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;    94WO-IB000207.
XX
PR   26-APR-1993;    93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.1%;  Score 401.6;  DB 2;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 1.2e-118;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||||||  | |  | | | | | | | | | | | | | | | | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
AAT74142
```

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.1%;  Score 401.6;  DB 2;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 1.2e-118;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |      | || |   || |  |   | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 3; Length 456;
Best Local Similarity 92.5%; Pred. No. 1.2e-118;
Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy    1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy   61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy  121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy  181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy  241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
          |||||||||||||||||||| | | |||| | || | || | | | | | | | | | |
Db  241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy  301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
          ||        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy  361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy  421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
          ||||||||||||||||||||||||||||||||||
Db  421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;

XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 82.5%; Score 376; DB 3; Length 456;
Best Local Similarity 89.0%; Pred. No. 2.1e-110;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy    1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy   61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy  121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy  181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy  241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
          | |||  |  |  | || |  | |  |  |  ||  |  | |
Db  241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy  301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
          ||        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy  361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy  421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
          ||||||||||||||||||||||||||||||||||
Db  421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

```
  Query Match           78.6%;  Score 358.4;  DB 3;  Length 456;
  Best Local Similarity  86.6%;  Pred. No. 9.5e-105;
  Matches  395;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| |||   |   ||||||     | |     | |   ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              | ||||||||||||||||||  |  |       | ||  |   ||  |  |    | |
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 7
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             77.2%;  Score 352;  DB 3;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 1.1e-102;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||     |   ||  |    | | ||  | |    ||  | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||  |  |        | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 3.6e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |      | ||  |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||  |||      ||   ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 3.6e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |      | ||  |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||  |           | | || | ||  |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 3.6e-102;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||  |  |       | ||  |    ||  |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 3.6e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |      | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||    | ||||  ||       |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match            76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 1.2e-101;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||  |  |      | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      || |||  |   |              |  |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64624
ID   AAC64624 standard; DNA; 456 BP.
XX
AC   AAC64624;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36348.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

  Query Match           76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 1.2e-101;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              |||||| |  ||  ||  | | | | |  |    |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||||  |  |  | ||  |  ||  |  |  | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match            60.8%;  Score 277.4;  DB 2;  Length 361;
  Best Local Similarity  89.3%;  Pred. No. 9.4e-79;
  Matches  299;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 303
              |||||||||||||||||  |  |       | || |    || |  |    | |  ||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 CATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match            60.8%;  Score 277.4;  DB 2;  Length 361;
  Best Local Similarity  89.3%;  Pred. No. 9.4e-79;
  Matches  299;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 303
              |||||||||||||||||  |  |       | || |    || |  |    | |  ||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 CATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 15, 2004, 04:14:40
```

Job time : 294.4 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:01:03 ; Search time 56.8 Seconds
(without alignments)
4455.240 Million cell updates/sec

Title: US-09-543-407-21
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	401.6	88.1	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	277.4	60.8	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
c	5	33.4	7.3	978	4	US-09-252-991A-86	Sequence 86, Appl
	6	33.4	7.3	1158	4	US-09-252-991A-96	Sequence 96, Appl
c	7	33.4	7.3	1206	4	US-09-252-991A-92	Sequence 92, Appl
	8	31.4	6.9	1059	4	US-09-252-991A-884	Sequence 884, App
c	9	31.4	6.9	1206	4	US-09-252-991A-801	Sequence 801, App
c	10	30.8	6.8	2831	4	US-09-477-135A-117	Sequence 117, App
	11	30.8	6.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	12	30.8	6.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	13	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	14	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	15	30.2	6.6	1178	2	US-08-107-676-26	Sequence 26, Appl
c	16	30.2	6.6	1178	4	US-09-295-820-26	Sequence 26, Appl
c	17	30.2	6.6	1211	2	US-08-107-676-2	Sequence 2, Appli
c	18	30.2	6.6	1211	4	US-09-295-820-2	Sequence 2, Appli
c	19	30.2	6.6	2052	1	US-08-443-104-5	Sequence 5, Appli
c	20	30.2	6.6	2052	1	US-08-238-130-6	Sequence 6, Appli
c	21	30.2	6.6	2052	1	US-08-442-859-5	Sequence 5, Appli
c	22	30.2	6.6	2052	2	US-08-398-489-5	Sequence 5, Appli
c	23	30.2	6.6	2052	2	US-08-894-772-1	Sequence 1, Appli
c	24	30.2	6.6	2052	2	US-09-207-844-1	Sequence 1, Appli
c	25	30.2	6.6	2052	4	US-09-252-509-1	Sequence 1, Appli
c	26	30.2	6.6	2052	5	PCT-US95-05534-5	Sequence 5, Appli
c	27	30	6.6	2325	4	US-09-252-991A-218	Sequence 218, App
	28	30	6.6	2784	4	US-09-252-991A-194	Sequence 194, App
c	29	30	6.6	2946	4	US-09-252-991A-227	Sequence 227, App
c	30	29.8	6.5	912	3	US-09-105-390-39	Sequence 39, Appl
c	31	29.8	6.5	993	3	US-09-105-390-55	Sequence 55, Appl
	32	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	33	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	34	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
c	35	29.8	6.5	1834	3	US-09-105-390-4	Sequence 4, Appli
c	36	29.6	6.5	615	4	US-09-291-922-17	Sequence 17, Appl
c	37	29.6	6.5	1782	4	US-09-252-991A-8703	Sequence 8703, Ap
c	38	29.6	6.5	1830	4	US-09-252-991A-8916	Sequence 8916, Ap
c	39	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	40	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	41	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	42	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	43	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
c	44	29.2	6.4	1215	4	US-09-489-039A-5440	Sequence 5440, Ap
	45	29.2	6.4	1245	1	US-09-337-913-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-788A-58

  Query Match             88.1%;  Score 401.6;  DB 1;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 3.4e-126;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 60.8%; Score 277.4; DB 1; Length 361;
Best Local Similarity 89.3%; Pred. No. 3.5e-84;
Matches 299; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 303
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 CATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA AND ITS USE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,927
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,108
; FILING DATE: 06-AUG-1977
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1581
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304

```
;       TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1344 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.19;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |  |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.19;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |  |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-252-991A-86/c
; Sequence 86, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 86
;   LENGTH: 978
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-86

  Query Match             7.3%;  Score 33.4;  DB 4;  Length 978;
  Best Local Similarity   62.7%;  Pred. No. 0.26;
  Matches   52;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy        232 GGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACC 291
              | ||||||||||||||||||   |||    |  | ||| | |    ||| |  | | |
Db        637 GACGCCGATGTAGGCCAGGGTCTTGATCGGCACCAGCAGGTCGGAGACCGCGGCGATGCG 578

Qy        292 CATGAAATGGCACATGCAGACCA 314
              | |||    ||  ||| ||||||
Db        577 CTTGATGCCGCCGATGAAGACCA 555


RESULT 6
US-09-252-991A-96
; Sequence 96, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 96
;   LENGTH: 1158
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-96

  Query Match             7.3%;  Score 33.4;  DB 4;  Length 1158;
  Best Local Similarity   62.7%;  Pred. No. 0.28;
  Matches   52;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy        232 GGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACC 291
              | ||||||||||||||||||   |||    |  | ||| | |    ||| |  | | |
Db         76 GACGCCGATGTAGGCCAGGGTCTTGATCGGCACCAGCAGGTCGGAGACCGCGGCGATGCG 135

Qy        292 CATGAAATGGCACATGCAGACCA 314
              | |||    ||  ||| ||||||
Db        136 CTTGATGCCGCCGATGAAGACCA 158


RESULT 7
US-09-252-991A-92/c
; Sequence 92, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 92
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-92

Query Match 7.3%; Score 33.4; DB 4; Length 1206;
Best Local Similarity 62.7%; Pred. No. 0.29;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 232 GGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACC 291
Db 651 GACGCCGATGTAGGCCAGGGTCTTGATCGGCACCAGCAGGTCGGAGACCGCGGCGATGCG 592

Qy 292 CATGAAATGGCACATGCAGACCA 314
Db 591 CTTGATGCCGCCGATGAAGACCA 569

RESULT 8
US-09-252-991A-884
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

Query Match 6.9%; Score 31.4; DB 4; Length 1059;
Best Local Similarity 43.3%; Pred. No. 1.3;
Matches 146; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

Qy 22 GCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGC 81
Db 97 GCGCTGGCAGCCGGCGTAGGCGATCATCGCGCCATTGTCGGTGCAGAAGCGCGGGCGGGC 156

Qy 82 GGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATT 141
Db 157 GTAGAACACCTGCCCCTTCATTTCGCCGAGCATCTTTTCCAGGCCGCTGCGCAGCGCCTG 216

Qy 142 TATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAA 201
Db 217 GTTGGCGCTGACACCGCCGGCGATCACCAGGTTCTTCAGGCCGGTCTGCTTCAAGGCGCG 276

Qy 202 ACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAAT 261
Db 277 ACGGCACTTGATCAGCAGGGTCTCGACCACCGCGGTCTGGAACGCCAGGGCGATGTCGCA 336

Qy 262 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAAC 321
Db 337 GCGGGTCTGCTCGCTGTCGTCGCCGGCCTCGACGCAACGCTGCCAGGTGTTCAGGGTAAA 396

Qy 322 GCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
Db 397 GGTCTTGAGCCCGCTGAAGCTGAAGTCCAGGCCGGGG 433

RESULT 9
US-09-252-991A-801/c
; Sequence 801, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 801
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-801

Query Match 6.9%; Score 31.4; DB 4; Length 1206;
Best Local Similarity 43.3%; Pred. No. 1.4;
Matches 146; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

Qy 22 GCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGC 81
Db 1128 GCGCTGGCAGCCGGCGTAGGCGATCATCGCGCCATTGTCGGTGCAGAAGCGCGGGCGGGC 1069

Qy 82 GGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATT 141
Db 1068 GTAGAACACCTGCCCCTTCATTTCGCCGAGCATCTTTTCCAGGCCGCTGCGCAGCGCCTG 1009

Qy 142 TATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAA 201
Db 1008 GTTGGCGCTGACACCGCCGGCGATCACCAGGTTCTTCAGGCCGGTCTGCTTCAAGGCGCG 949

Qy 202 ACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAAT 261
Db 948 ACGGCACTTGATCAGCAGGGTCTCGACCACCGCGGTCTGGAACGCCAGGGCGATGTCGCA 889

Qy 262 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAAC 321
Db 888 GCGGGTCTGCTCGCTGTCGTCGCCGGCCTCGACGCAACGCTGCCAGGTGTTCAGGGTAAA 829

Qy 322 GCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
Db 828 GGTCTTGAGCCCGCTGAAGCTGAAGTCCAGGCCGGGG 792

RESULT 10
US-09-477-135A-117/c
; Sequence 117, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 117
;   LENGTH: 2831
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-117

  Query Match             6.8%;  Score 30.8;  DB 4;  Length 2831;
  Best Local Similarity   52.3%;  Pred. No. 3.4;
  Matches   68;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | | | || ||     | |  |||| || | ||||| |  | || ||  |  || ||
Db        840 GTTCAGGCCGCCGAACCCCAACAAACCGCTACCCGTGAGCCCGATACCTCGGTTGCCGTC 781

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTT 276
              ||||  ||   | | ||  | ||| |     ||||  |  ||    |||   | | |
Db        780 TCCGGTATTGCCGAAGCCGATGTTGTTGCTGCCGGTGTTGCCAAACCCGATGTTGTTGCT 721

Qy        277 ACCCGTGTTG 286
              || ||||||
Db        720 GCCGGTGTTG 711


RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.8%;  Score 30.8;  DB 3;  Length 4403765;
  Best Local Similarity   52.3%;  Pred. No. 89;
  Matches   68;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | | | || ||     | |  |||| || | ||||| |  | || ||  |  || ||
Db     433864 GTTCAGGCCGCCGAACCCCAACAAACCGCTACCCGTGAGCCCGATACCTCGGTTGCCGTC 433923

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTT 276
              ||||  ||   | | ||  | ||| |     ||||  |  ||    |||   | | |
Db     433924 TCCGGTATTGCCGAAGCCGATGTTGTTGCTGCCGGTGTTGCCAAACCCGATGTTGTTGCT 433983

Qy        277 ACCCGTGTTG 286
              || ||||||
Db     433984 GCCGGTGTTG 433993


RESULT 12
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             6.8%;  Score 30.8;  DB 3;  Length 4411529;
  Best Local Similarity   52.3%;  Pred. No. 89;
  Matches   68;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | | | || ||     | |  |||| || | ||||| |  | || ||  |  || ||
Db     433781 GTTCAGGCCGCCGAACCCCAACAAACCGCTACCCGTGAGCCCGATACCTCGGTTGCCGTC 433840

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTT 276
              ||||  ||   | | ||  | ||| |     ||||  |  ||    |||   | | |
Db     433841 TCCGGTATTGCCGAAGCCGATGTTGTTGCTGCCGGTGTTGCCAAACCCGATGTTGTTGCT 433900

Qy        277 ACCCGTGTTG 286
              || ||||||
Db     433901 GCCGGTGTTG 433910


RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4403765;
  Best Local Similarity   51.9%;  Pred. No. 95;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || ||||||| || ||  |  ||   | |||||| |||| |  |   || ||
Db    1357218 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357159

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
              ||     | | |   ||  | ||| | ||   |   ||   ||  || | |  |||
Db    1357158 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357099

Qy        165 TGCGCTTGCTCTG 177
              ||   |  |||
```

```
Db    1357098 CCCGGACGTGCTG 1357086


RESULT 14
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4411529;
  Best Local Similarity   51.9%;  Pred. No. 95;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || |||||||| || ||  |  ||   | |||||| |||| |  |   || ||
Db    1357750 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357691

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
              ||     | | |    || | ||| | ||   |  ||   || || | |  |||
Db    1357690 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357631

Qy        165 TGCGCTTGCTCTG 177
              ||   |  |||
Db    1357630 CCCGGACGTGCTG 1357618


RESULT 15
US-08-107-676-26/c
; Sequence 26, Application US/08107676
; Patent No. 5955356
;  GENERAL INFORMATION:
;     APPLICANT:  Content, Jean
;     APPLICANT:  DeWit, Lucas
;     APPLICANT:  DeBruyn, Jacqueline
;     TITLE OF INVENTION:  Mycobacterium Polypeptides and Nucleic
;     TITLE OF INVENTION:  Acids Encoding Them for Diagnosis and Control of
;     TITLE OF INVENTION:  Tuberculosis
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Merchant & Gould
;       STREET:  3100 No. 5955356west Center
;       CITY:  Minneapolis
;       STATE:  MN
;       COUNTRY:  USA
;       ZIP:  55402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/107,676
;       FILING DATE:  04-OCT-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/EP92/00268
;       FILING DATE:  07-FEB-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  EP 91400388.4
;       FILING DATE:  14-FEB-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kowalchyk, Katherine M.
;       REGISTRATION NUMBER:  36,848
;       REFERENCE/DOCKET NUMBER:  8076.89USWO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  612-332-5300
;       TELEFAX:  612-332-9081
;  INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1178 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycobacterium tuberculosis
;     IMMEDIATE SOURCE:
;       CLONE:  Antigen 85C from M. tuberculosis
US-08-107-676-26

  Query Match             6.6%;  Score 30.2;  DB 2;  Length 1178;
  Best Local Similarity   55.1%;  Pred. No. 3.4;
  Matches   59;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        218 GCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTA 277
              || | |||| |  || |||||| || |    | |||| | ||      |  || | | |
Db        958 GCCGGTATGTTGTCGCCGCCGAGGTCGCTGGGTGTGCCGTTACCGCAGTACACCCAGATC 899

Qy        278 CCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAACGCT 324
              |  |||||||  | |  |   ||   ||    |||| |||  || |
Db        898 CGGGTGTTGTTGGCGACCAGGCGGGGAATCTGAACCATTGGGTCGTT 852


Search completed: March 15, 2004, 10:14:33
Job time : 80.8 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 07:24:24 ; Search time 340.6 Seconds
(without alignments)
4928.974 Million cell updates/sec

Title: US-09-543-407-21
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	35.4	7.8	494	9	US-09-864-761-11428	Sequence 11428, A
c 2	35	7.7	616	12	US-10-424-599-88945	Sequence 88945, A
c 3	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c 4	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c 5	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c 6	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c 7	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
c 8	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
9	33.2	7.3	2329	15	US-10-260-238-1191	Sequence 1191, Ap
10	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
c 11	33	7.2	1350	15	US-10-369-493-41653	Sequence 41653, A
12	32.8	7.2	2530	15	US-10-439-741-5	Sequence 5, Appli
13	32.4	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
14	32.4	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c 15	32.4	7.1	1842	14	US-10-359-050-5	Sequence 5, Appli
c 16	32.4	7.1	1863	14	US-10-359-050-6	Sequence 6, Appli
c 17	32.4	7.1	1885	14	US-10-359-050-8	Sequence 8, Appli
c 18	32.4	7.1	5741	14	US-10-359-050-10	Sequence 10, Appl
c 19	32.4	7.1	12645	14	US-10-359-050-13	Sequence 13, Appl
c 20	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
c 21	31.8	7.0	843	15	US-10-398-221-1629	Sequence 1629, Ap
22	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
23	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
24	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c 25	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
26	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
27	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
28	31.6	6.9	3532	14	US-10-017-161-1913	Sequence 1913, Ap
29	31.6	6.9	3532	15	US-10-292-798-1569	Sequence 1569, Ap
c 30	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
c 31	31.4	6.9	897	12	US-10-425-114-2086	Sequence 2086, Ap
c 32	31.4	6.9	1026	9	US-09-815-242-7701	Sequence 7701, Ap
c 33	31.4	6.9	1026	12	US-10-282-122A-30016	Sequence 30016, A
c 34	31.4	6.9	1268	12	US-10-425-114-26408	Sequence 26408, A
c 35	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
c 36	31.2	6.8	6607	12	US-10-359-120-8	Sequence 8, Appli
c 37	31.2	6.8	6787	12	US-10-359-120-7	Sequence 7, Appli
c 38	31.2	6.8	6859	12	US-10-359-120-10	Sequence 10, Appl
c 39	31.2	6.8	6970	12	US-10-359-120-6	Sequence 6, Appli
c 40	31.2	6.8	7003	12	US-10-359-120-73	Sequence 73, Appl
c 41	31.2	6.8	7042	12	US-10-359-120-9	Sequence 9, Appli
c 42	31.2	6.8	7186	12	US-10-359-120-72	Sequence 72, Appl
c 43	31.2	6.8	7315	12	US-10-359-120-77	Sequence 77, Appl
c 44	31.2	6.8	7351	12	US-10-359-120-4	Sequence 4, Appli
c 45	31.2	6.8	7498	12	US-10-359-120-76	Sequence 76, Appl

ALIGNMENTS

RESULT 1
US-09-864-761-11428/c
; Sequence 11428, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11428
;   LENGTH: 494
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005865.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.97
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
US-09-864-761-11428

  Query Match             7.8%;  Score 35.4;  DB 9;  Length 494;
  Best Local Similarity   53.2%;  Pred. No. 0.13;
  Matches   75;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy        267 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAACGCTAA 326
              ||  ||||  |  |  || | || |||||||     ||  ||  | || || | |  ||
Db        267 TCTTCTGGGGAGACTCATTCTCACACATGAAATACACCAGACACCCAAGAGGCATGTGAA 208

Qy        327 AAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGAC 386
              |||  |  || || |  |  |  |  ||    || |  ||| ||   ||||  || |||
Db        207 AAAGTGAGACATGATGGGTGAACCCTAAAAAGGCATGAACTCCATTCTGGAAAAACAGGG 148

Qy        387 CGCATCTGATTCCAGCGTAAT 407
              |  ||  |  ||  |  |  | |
Db        147 CTGATGGGGTTGGATCAGAGT 127


RESULT 2
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
;   LENGTH: 616
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(616)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

  Query Match             7.7%;  Score 35;  DB 12;  Length 616;
  Best Local Similarity   50.3%;  Pred. No. 0.2;
  Matches   86;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
              | | |  |||  ||  || ||| || || |||  | |  |  | | | | | |  | |
Db        556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497

Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              | | ||    || |  | |    || ||  |   | | |||    ||  |  | | |
Db        496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437

Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGC 271
              || ||||||  |  |   ||| |||  ||||| ||| ||      | ||||
Db        436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTACGGTGAAGCAGC 386


RESULT 3
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Qiu, Dewen
;  APPLICANT:  Remick, Dean
;  TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
;  TITLE OF INVENTION:  RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
;  TITLE OF INVENTION:  DESICCATION
;  FILE REFERENCE: 21829/71
;  CURRENT APPLICATION NUMBER: US/09/835,684
;  CURRENT FILING DATE:  2001-04-16
;  PRIOR APPLICATION NUMBER: 60/198,359
;  PRIOR FILING DATE: 2000-04-19
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-835-684-6

  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.65;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |     ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  DeRocher, Jay
;  TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE: 21829/91
;  CURRENT APPLICATION NUMBER: US/09/880,371
;  CURRENT FILING DATE:  2001-06-13
;  PRIOR APPLICATION NUMBER: 60/211,585
;  PRIOR FILING DATE: 2000-06-15
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
```

; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.65;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 5
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.65;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 6
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-010-390-6

Query Match 7.5%; Score 34; DB 14; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.65;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 7
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (EBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-441-736-5

Query Match 7.5%; Score 34; DB 15; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.65;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 8
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761

; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1662)
US-10-156-761-3483

Query Match 7.3%; Score 33.2; DB 14; Length 1662;
Best Local Similarity 52.1%; Pred. No. 1.4;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
Db 590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531

Qy 110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
Db 530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

Qy 170 TTGCTCTGCAAAGCGATGCCCG 191
Db 470 TTGAACTTCTTGGCGGCGACGG 449

RESULT 9
US-10-260-238-1191
; Sequence 1191, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1191
; LENGTH: 2329
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (221)..(221)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (1565)..(1565)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-1191

Query Match 7.3%; Score 33.2; DB 15; Length 2329;
Best Local Similarity 51.3%; Pred. No. 1.6;
Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 286 GTTACCCATGAAATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTC 345
Db 425 GCTATGCATGAGATGGGTCTTGCATCCCGTAGACCATTCAGAAAATGTGCAAGGGTTGTC 484

Qy 346 GGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTA 405
Db 485 GGTGAGGTACTAGGTAAGTTCCATCCACATGGAGATTCGGCCGTATATGAAACCTTAGTA 544

Qy 406 ATGGTGCGTCAGGTTGGTTTTGGCAACAAC 435
Db 545 CGGATGGCTCAGGTTGACTTCGTCCCAAAC 574

RESULT 10
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 7.3%; Score 33.2; DB 14; Length 9025608;
Best Local Similarity 52.1%; Pred. No. 77;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
Db 4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

Qy 110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
Db 4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

Qy 170 TTGCTCTGCAAAGCGATGCCCG 191

```
              ||| || |   ||| | | |
Db    4331901 TTGAACTTCTTGGCGGCGACGG 4331922


RESULT 11
US-10-369-493-41653/c
; Sequence 41653, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 41653
;   LENGTH: 1350
;   TYPE: DNA
;   ORGANISM: SPHINGOMONAS
US-10-369-493-41653

  Query Match             7.2%;  Score 33;  DB 15;  Length 1350;
  Best Local Similarity   60.7%;  Pred. No. 1.4;
  Matches   54;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy        234 CGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCA 293
              |||| |||  ||  |||||  |||    |  | |||||  ||   | |||||||| | |
Db        955 CGCCCATGGTGGAGAGGGTCAGGAAGCTCTCGGTCAGCATGTCGGCGGTGTTGTTGTCGA 896

Qy        294 TGAAATGGCACATGCAGACCAGTGGAACG 322
              | |   | | |||| || | | ||| ||
Db        895 TCACCGGACCCATGAAGGGCTGGGGATCG 867


RESULT 12
US-10-439-741-5
; Sequence 5, Application US/10439741
; Publication No. US20030235575A1
; GENERAL INFORMATION:
;  APPLICANT: Matzuk, Martin M
;  APPLICANT:  Wei, Yan
;  APPLICANT:  Yuchen, Bai
;  TITLE OF INVENTION: Oligoadenylate Synthetase-Like Genes
;  FILE REFERENCE: P02432US1/10203964
;  CURRENT APPLICATION NUMBER: US/10/439,741
;  CURRENT FILING DATE:  2003-05-16
;  PRIOR APPLICATION NUMBER: US 60/381,408
;  PRIOR FILING DATE: 2002-05-17
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 2530
;   TYPE: DNA
;   ORGANISM: Mouse
US-10-439-741-5

  Query Match             7.2%;  Score 32.8;  DB 15;  Length 2530;
  Best Local Similarity   48.9%;  Pred. No. 2.3;
  Matches   88;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        247 CAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 306
              |||   | ||   | | | ||  || |    ||  || ||      ||||| ||  |
Db       1300 CAGCAGGAGGTGGAATTTCATATGCAGACAGCCTATGATGCCCTGTATGAAGTGAGAAGA 1359

Qy        307 GCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
                |||  |   |||  | ||| || | | |       || || |  |  | || |   |

Db       1360 AAAGAGAATCATAACTGTGAAATCTACAACAAAGTATACGCCCGACTCATCCGTGAGTGC 1419

Qy        367 GCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTT 426
              ||  ||||| ||  ||  ||  |   |  | || | |||| || ||||   | |||
Db       1420 ACCATGCTGGGCAAGGAGGGCGAGTTCAACATCTGCTTCATGGAGCTTCAGCAAGACTTT 1479


RESULT 13
US-10-333-631-1
; Sequence 1, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer AG
;  TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE (
;  TITLE OF INVENTION:  ENZYME
;  FILE REFERENCE: LIO121 Foreign Countries
;  CURRENT APPLICATION NUMBER: US/10/333,631
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US 60/221,917
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/280,112
;  PRIOR FILING DATE: 2001-04-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 568
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90)..(90)
;   OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

  Query Match             7.1%;  Score 32.4;  DB 14;  Length 568;
  Best Local Similarity   46.8%;  Pred. No. 1.5;
  Matches  102;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCC 214
              | ||| |    |  ||||| |||||   ||  || ||     ||||| | |  |  ||
Db        103 CTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCTGAAATGGTCCGTCACC 162

Qy        215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGG 274
              ||  |    || || | || | || | ||| ||||||||     |     | |
Db        163 CAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTACCTGGATCTGCGGATA 222

Qy        275 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCG 334
              | | || ||    |  | |  |  || ||||          | | |  | ||  ||
Db        223 GCCCACATGCTGGAGGGCTCGGAGAAGAACCTGCACTTTGTCCATATGGTGTACACAACG 282

Qy        335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCG 372
                |  ||  || ||       ||| |||  ||  | |
Db        283 GCGCTGGTGGAGGACACACTCACGGAAATCTCGGAGTG 320


RESULT 14
US-10-333-631-4
; Sequence 4, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer AG
;  TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE (
;  TITLE OF INVENTION:  ENZYME
;  FILE REFERENCE: LIO121 Foreign Countries
;  CURRENT APPLICATION NUMBER: US/10/333,631
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US 60/221,917
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/280,112
;  PRIOR FILING DATE: 2001-04-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
```

; LENGTH: 972
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-333-631-4

Query Match 7.1%; Score 32.4; DB 14; Length 972;
Best Local Similarity 46.8%; Pred. No. 2;
Matches 102; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```
Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCC 214
              | |||| |    |  ||||| |||||   ||   || ||     |||||| |  |  |  ||
Db        202 CTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCTGAAATGGTCCGTCACC 261

Qy        215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGG 274
              ||  |    || || | || | || | ||| ||||||||    |      |  |
Db        262 CAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTACCTGGACCTGCGGATA 321

Qy        275 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCG 334
              | | || ||      |  | |  | || ||||         | | |  | ||  ||
Db        322 GCCCACATGCTGGAGGGCTCGGAGAAGAACCTGCACTTTGTCCATATGGTGTACACAACG 381

Qy        335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCG 372
              |  ||  || ||      ||| |||  ||  | |
Db        382 GCGCTGGTGGAGGACACACTCACGGAAATCTCGGAGTG 419
```

RESULT 15
US-10-359-050-5/c
; Sequence 5, Application US/10359050
; Publication No. US20030186291A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS DEUTSCHLAND GMBH
; TITLE OF INVENTION: GENETICALLY ENGINEERED PHIC31-INTEGRASE GENES
; FILE REFERENCE: AR03-001
; CURRENT APPLICATION NUMBER: US/10/359,050
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US 60/354,741
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 1842
; TYPE: DNA
; ORGANISM: C-31-Int-CO
US-10-359-050-5

Query Match 7.1%; Score 32.4; DB 14; Length 1842;
Best Local Similarity 55.3%; Pred. No. 2.7;
Matches 63; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATC 268
              || | |||  |||| ||| |  ||| |  ||||  |||| || ||| |    | | |  |
Db        770 TTGCACAGGCCGGTGATGCTGCCGGGGTGGATGGCGGCCTGGCTGCCGGGCTTGAAGGGC 711

Qy        269 AGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAACG 322
              || || ||    || ||| |  || |   |  |||  ||   ||   || ||
Db        710 AGGTGCTTGTGGGTCTTGATCTCCCTCCACCACCACCTGATCACGTCGGGCTCG 657
```

Search completed: March 15, 2004, 14:16:06
Job time : 353.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:52:04 ; Search time 2500.2 Seconds
(without alignments)
6417.435 Million cell updates/sec

Title: US-09-543-407-21
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-21	Sequence 21, Appl
2	401.6	88.1	456	6	US-08-233-642A-56	Sequence 56, Appl
3	401.6	88.1	456	23	US-09-543-407-1	Sequence 1, Appli
4	376	82.5	456	23	US-09-543-407-27	Sequence 27, Appl
5	358.4	78.6	456	23	US-09-543-407-19	Sequence 19, Appl
6	352	77.2	456	23	US-09-543-407-17	Sequence 17, Appl
7	350.4	76.8	456	23	US-09-543-407-11	Sequence 11, Appl
8	350.4	76.8	456	23	US-09-543-407-13	Sequence 13, Appl
9	350.4	76.8	456	23	US-09-543-407-23	Sequence 23, Appl
10	350.4	76.8	456	23	US-09-543-407-25	Sequence 25, Appl
11	348.8	76.5	456	23	US-09-543-407-15	Sequence 15, Appl
12	348.8	76.5	456	23	US-09-543-407-29	Sequence 29, Appl
13	277.4	60.8	361	6	US-08-233-642A-54	Sequence 54, Appl
14	230.4	50.5	456	23	US-09-543-407-3	Sequence 3, Appli
15	230.4	50.5	675	13	US-08-978-878-3	Sequence 3, Appli
16	230.4	50.5	675	31	US-09-741-873B-3	Sequence 3, Appli
17	228	50.0	453	92	US-60-352-946-1	Sequence 1, Appl
18	228	50.0	453	101	US-60-444-371-1	Sequence 1, Appli
19	217.6	47.7	477	17	US-09-252-691-172	Sequence 172, App
20	217.6	47.7	477	17	US-09-252-691C-172	Sequence 172, App
21	217.6	47.7	477	51	US-10-417-886-172	Sequence 172, App
22	186.4	40.9	396	13	US-08-978-878-1	Sequence 1, Appli
23	161.6	35.4	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
28	50.8	11.1	78	23	US-09-543-407-53	Sequence 53, Appl
c 29	50.6	11.1	78	23	US-09-543-407-52	Sequence 52, Appl
c 30	49.4	10.8	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
31	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
32	41	9.0	78	23	US-09-543-407-49	Sequence 49, Appl
c 33	40.8	8.9	78	23	US-09-543-407-48	Sequence 48, Appl
34	38	8.3	1947	1	PCT-US03-04154-49	Sequence 49, Appl
35	38	8.3	1947	92	US-60-359-557-49	Sequence 49, Appl
36	36.8	8.1	1304	51	US-10-437-963-19773	Sequence 19773, A
c 37	35.4	7.8	494	1	PCT-US01-00663-7203	Sequence 7203, Ap
c 38	35.4	7.8	494	35	US-09-864-761-11428	Sequence 11428, A
c 39	35.4	7.8	494	46	US-10-182-993-6957	Sequence 6957, Ap
c 40	35.4	7.8	494	47	US-10-203-134-7259	Sequence 7259, Ap
c 41	35.4	7.8	494	47	US-10-203-135-7489	Sequence 7489, Ap
c 42	35.4	7.8	494	47	US-10-203-136-7411	Sequence 7411, Ap
c 43	35.4	7.8	494	47	US-10-203-137-7203	Sequence 7203, Ap
c 44	35.4	7.8	494	47	US-10-203-139-7118	Sequence 7118, Ap
45	35	7.7	100	45	US-10-146-492B-78	Sequence 78, Appl

ALIGNMENTS

```
RESULT 1
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;  LENGTH: 456
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;  OTHER INFORMATION: sequence containing the replacement fragment
;  OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match                 100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity       100.0%;  Pred. No. 2.6e-132;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.1%;  Score 401.6;  DB 6;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 3.4e-115;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |      | ||  |    ||  |   |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.1%;  Score 401.6;  DB 23;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 3.4e-115;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |      | ||  |    ||  |   |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

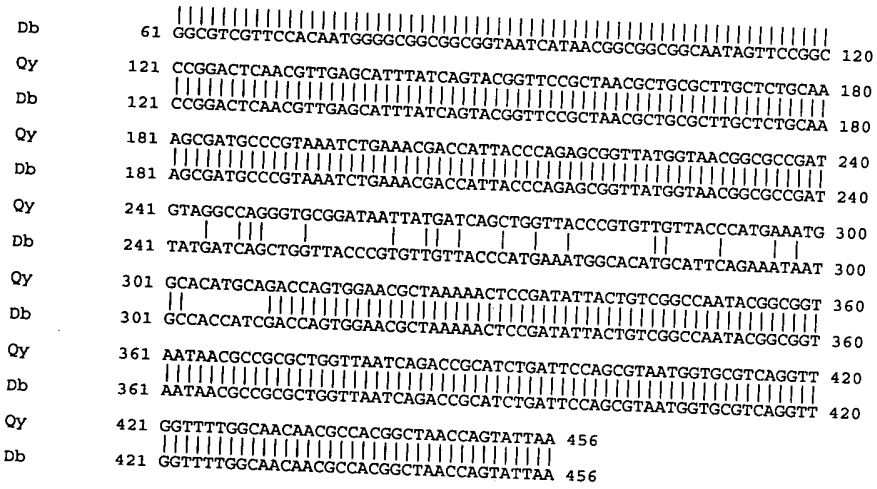
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             82.5%;  Score 376;  DB 23;  Length 456;
  Best Local Similarity   89.0%;  Pred. No. 3.9e-107;
  Matches  406;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

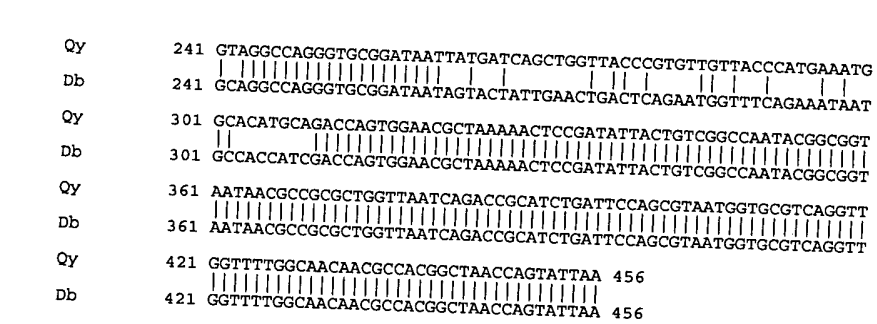
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```



RESULT 5
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

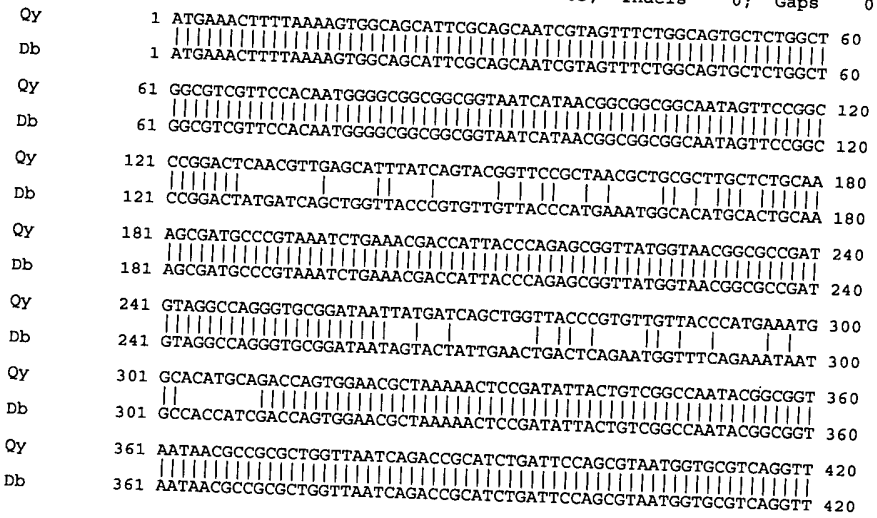
Query Match 78.6%; Score 358.4; DB 23; Length 456;
Best Local Similarity 86.6%; Pred. No. 1.4e-101;
Matches 395; Conservative 0; Mismatches 61; Indels 0; Gaps 0;





RESULT 6
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.7%; Pred. No. 1.4e-99;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;



```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 4.5e-99;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |        | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||| |          | | || | ||   |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 4.5e-99;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |        | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||| |          | | || | ||   |    | | |  |
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match           76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 4.5e-99;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||||  |  |     | |||  ||  |  |  | ||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||     ||||||||||||||||||||  |||     ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |   ||  |  |  ||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match           76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 4.5e-99;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |           ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              ||||||||||||||||||  |  |     | || |  || |  | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match           76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 1.4e-98;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |  ||  ||    | | | |  | |    |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||||  |  |      | ||  |   ||  |   |  | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 76.5%; Score 348.8; DB 23; Length 456;
Best Local Similarity 85.3%; Pred. No. 1.4e-98;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              |||||||||||||||||||  |  |      | ||  |   ||  |   |  | |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 60.8%; Score 277.4; DB 6; Length 361;
Best Local Similarity 89.3%; Pred. No. 3.9e-76;
Matches 299; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |||||  |||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 303
              |||||||||||||||||  | |  |  | ||| |  || | |  |  | | ||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 CATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-3

  Query Match           50.5%;  Score 230.4;  DB 23;  Length 456;
  Best Local Similarity  69.1%;  Pred. No. 2.6e-61;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||| || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | ||||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              || || |||||  | ||| |                | ||||   ||    | | | |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||    |  || |||||||||| |||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
;  APPLICANT: NORMARK, Staffan
;  APPLICANT: OLSEN, Arne
;  TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
;  FILE REFERENCE: 012889-081
;  CURRENT APPLICATION NUMBER: US/08/978,878
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: SE 8801723-1
;  EARLIER FILING DATE: 1988-05-06
;  EARLIER APPLICATION NUMBER: US 07/347,189
;  EARLIER FILING DATE: 1989-05-04
;  EARLIER APPLICATION NUMBER: US 07/789,437
;  EARLIER FILING DATE: 1991-11-06
;  EARLIER APPLICATION NUMBER: US 07/970,846
;  EARLIER FILING DATE: 1992-11-03
;  EARLIER APPLICATION NUMBER: US 08/187,865
;  EARLIER FILING DATE: 1994-01-28
;  EARLIER APPLICATION NUMBER: US 08/318,519
;  EARLIER FILING DATE: 1994-10-05
;  EARLIER APPLICATION NUMBER: US 08/495,959
;  EARLIER FILING DATE: 1995-06-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-08-978-878-3

  Query Match           50.5%;  Score 230.4;  DB 13;  Length 675;
  Best Local Similarity  69.1%;  Pred. No. 2.9e-61;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||| || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | ||||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              || || |||||  | ||| |                | ||||   ||    | | | |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||    |  || |||||||||| |||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

Search completed: March 15, 2004, 13:42:47
Job time : 2503.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 04:16:29 ; Search time 53.4 Seconds
(without alignments)
2945.701 Million cell updates/sec

Title: US-09-543-407-21
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 628836 seqs, 172478556 residues

Total number of hits satisfying chosen parameters: 1257672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	230.4	50.5	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	186.4	40.9	396	5	US-09-741-873C-1	Sequence 1, Appli
c	3	33.4	7.3	532	6	US-10-767-701-538	Sequence 538, App
c	4	31.4	6.9	587	6	US-10-767-701-22392	Sequence 22392, A
c	5	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
	6	30.2	6.6	519599	6	US-10-765-790-73	Sequence 73, Appl
	7	30	6.6	1311	1	PCT-US03-19794-17	Sequence 17, Appl
	8	30	6.6	1533	1	PCT-US03-19794-5	Sequence 5, Appli
	9	29.8	6.5	1812	6	US-10-767-701-14980	Sequence 14980, A
	10	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
	11	29.4	6.4	1355	6	US-10-045-674A-593	Sequence 593, App
c	12	29.2	6.4	531	6	US-10-767-701-7332	Sequence 7332, Ap
c	13	29.2	6.4	290892	6	US-10-767-471-10809	Sequence 10809, A
c	14	29	6.4	1682	6	US-10-781-014-555	Sequence 555, App
c	15	29	6.4	1694	6	US-10-781-014-553	Sequence 553, App
c	16	28.8	6.3	165097	6	US-10-765-790-77	Sequence 77, Appl
	17	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
c	18	28.6	6.3	1194	1	PCT-US04-04914-4	Sequence 4, Appli
c	19	28.6	6.3	2229	1	PCT-US04-04914-18	Sequence 18, Appl
	20	28.4	6.2	1005	6	US-10-417-884A-3570	Sequence 3570, Ap
	21	28.4	6.2	1126118	6	US-10-767-471-10615	Sequence 10615, A
	22	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
	23	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	24	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	25	28.2	6.2	1633	6	US-10-767-701-14878	Sequence 14878, A
c	26	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
	27	28.2	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	28	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
	29	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
	30	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	31	27.6	6.1	2830	6	US-10-416-330-16	Sequence 16, Appl
	32	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	33	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	34	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	35	27.4	6.0	716	6	US-10-767-701-4864	Sequence 4864, Ap
c	36	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
c	37	27.4	6.0	1989	6	US-10-486-316-3	Sequence 3, Appli
c	38	27.4	6.0	1989	6	US-10-486-316-5	Sequence 5, Appli
	39	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
c	40	27.2	6.0	410	6	US-10-767-701-20998	Sequence 20998, A
	41	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
c	42	27.2	6.0	691	6	US-10-767-701-6541	Sequence 6541, Ap
c	43	27.2	6.0	2880	6	US-10-100-683-5057	Sequence 5057, Ap
c	44	27.2	6.0	2881	6	US-10-100-683-2682	Sequence 2682, Ap
c	45	27.2	6.0	7380	6	US-10-417-375A-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 50.5%; Score 230.4; DB 5; Length 675;
Best Local Similarity 69.1%; Pred. No. 4.8e-69;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 40.9%; Score 186.4; DB 5; Length 396;
Best Local Similarity 66.9%; Pred. No. 4.1e-54;
Matches 265; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-10-767-701-538/c
; Sequence 538, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 538
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS117067_1
US-10-767-701-538

Query Match 7.3%; Score 33.4; DB 6; Length 532;
Best Local Similarity 51.0%; Pred. No. 0.086;
Matches 79; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
Qy        218 GCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTA 277

Db        317 GCGGCGAAGGAGCCGATGAGGGAGTAGACGTTGAGGATGCCCATCAGGATCTCCAGCTGC 258

Qy        278 CCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGATA 337

Db        257 ACGTCGGTGATTCCCAGGTCCTTCTTGATGTATAGCGACGCACCGCTCATCACTCCGATA 198

Qy        338 TTACTGTCGGCCAATACGGCGGTAATAACGCCGCG 372

Db        197 TCATAGCCAAGAAGGATGGAGGTCATGGAGGCGAG 163
```

RESULT 4
US-10-767-701-22392/c
; Sequence 22392, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 22392
; LENGTH: 587
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 14188348
US-10-767-701-22392

Query Match 6.9%; Score 31.4; DB 6; Length 587;
Best Local Similarity 53.7%; Pred. No. 0.43;
Matches 65; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTA 289
              | | ||||||||||| | | |||     | | |    | || |    || || | ||
Db        300 AAGACGCCGATGTAGCGCCGCGTGTTCTCAGTGACACGCTGCAAGAACTCCTCGTCGATA 241

Qy        290 CCCATGAAATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              || |  ||||| |||||  | ||||     |      |   | |  || || || | |||
Db        240 TCCTTATAATGGAACATGTCGTCCAGCTCGATCTGGACCGCGCGTATCTTCCTATTGGCG 181

Qy        350 A 350
              |
Db        180 A 180
```

RESULT 5
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.1;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
              |||   | | | | || ||   | | || |  || | |||    || |||| ||||||||
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
              |  | |   | |  |||  |  |   |   |||  |  ||||||||
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583
```

RESULT 6
US-10-765-790-73
; Sequence 73, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73
; LENGTH: 519599
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-73

Query Match 6.6%; Score 30.2; DB 6; Length 519599;
Best Local Similarity 48.9%; Pred. No. 23;
Matches 111; Conservative 0; Mismatches 113; Indels 3; Gaps 1;

```
Qy         85 GGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTAT 144
              |||  | || |  || |   |    | |   | |||| || | |   | ||  ||| |||
Db     207952 GGCTCTTATGAACACTGGTACCAGCAGAAGCCAGGCCAGGCCCCCGTGCTGGTCATCTAT 208011

Qy        145 CAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACG 204
              |   | ||  |      | |   | |  ||  || | |  | || |   |||||  |  |
Db     208012 GATAGCAGTGACCGGCCCTCAAGGATCCCTGAGCGATTCTCTGGCTCCAAATCAGGCAAC 208071

Qy    0   205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTAT 264
              || |  |||| || |     |||   | || || | |  ||| ||  || | |||
Db    .208072 ACAACCACCCTGACCATCACTGGGGCCCAGGCTGAGGATG---AGGCTGATTATTACTAT 208128

Qy        265 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGA 311
              |   | | |  | || || | |   |    | || |||| |||||
Db     208129 CAGTTGATAGACAACCATGCTACTCAACTCACAGTGACACAGGCAGA 208175
```

RESULT 7
PCT-US03-19794-17
; Sequence 17, Application PC/TUS0319794
; GENERAL INFORMATION:
; APPLICANT: Divergence, Inc.
; TITLE OF INVENTION: NEMATODE PHOSPHOETHANOLAMINE
; TITLE OF INVENTION: N-METHYLTRANSFERASE-LIKE SEQUENCES
; FILE REFERENCE: 12557-011WO1
; CURRENT APPLICATION NUMBER: PCT/US03/19794
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: US 60/390,490
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: Ascaris suum
PCT-US03-19794-17

Query Match 6.6%; Score 30; DB 1; Length 1311;
Best Local Similarity 51.5%; Pred. No. 1.9;
Matches 69; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        156 CGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCA 215
              ||  |||   |   ||     || |  || ||  |  ||| | ||    |  ||   | |
Db        636 CGAAAACTTAGAAGTTCTGAAGCGATTCGGTGATCTAAAACCGGATTGTAAAATGCTCGA 695

Qy        216 GAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGT 275
              | ||||  ||| | ||| |  | ||   |||||| |||     | |  |    ||||||
Db        696 CATCGGTGTTGGGATCGGTGGAGGTGCCCGCCAGGCTGCTAGGGAATTCGGAGCGCTGGT 755

Qy        276 TACCCGTGTTGTTA 289
              |  | ||  | | ||
Db        756 TCTCGGTATGGATA 769
```

RESULT 8
PCT-US03-19794-5
; Sequence 5, Application PC/TUS0319794
; GENERAL INFORMATION:
; APPLICANT: Divergence, Inc.
; TITLE OF INVENTION: NEMATODE PHOSPHOETHANOLAMINE
; TITLE OF INVENTION: N-METHYLTRANSFERASE-LIKE SEQUENCES
; FILE REFERENCE: 12557-011WO1
; CURRENT APPLICATION NUMBER: PCT/US03/19794
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: US 60/390,490

```
;  PRIOR FILING DATE: 2002-06-21
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 1533
;   TYPE: DNA
;   ORGANISM: Ascaris suum
PCT-US03-19794-5

  Query Match             6.6%;  Score 30;  DB 1;  Length 1533;
  Best Local Similarity   51.5%;  Pred. No. 2;
  Matches   69;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        156 CGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCA 215
              ||  |||   |   ||     || |  || ||  |  ||| | ||    |  ||   | |
Db        695 CGAAAACTTAGAAGTTCTGAAGCGATTCGGTGATCTAAAACCGGATTGTAAAATGCTCGA 754

Qy        216 GAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGT 275
              | ||||  ||| | ||| |  | ||    ||||| |||     | |  |    ||||||
Db        755 CATCGGTGTTGGGATCGGTGGAGGTGCCCGCCAGGCTGCTAGGGAATTCGGAGCGCTGGT 814

Qy        276 TACCCGTGTTGTTA 289
              |  | || | | ||
Db        815 TCTCGGTATGGATA 828


RESULT 9
US-10-767-701-14980
; Sequence 14980, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14980
;   LENGTH: 1812
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS5075_1
US-10-767-701-14980

  Query Match             6.5%;  Score 29.8;  DB 6;  Length 1812;
  Best Local Similarity   51.1%;  Pred. No. 2.5;
  Matches   70;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAA 297
              ||  |  | | |||||||  ||| || | |||  |     |    ||  |    | ||||
Db       1207 GAAGCCGCCAAGGGTGGTGATGATGAGGATGACGACCTAGAGGATGAGGAAGACGATGAG 1266

Qy        298 ATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              | |||| | |  ||| ||    || || |   |  ||||  | | || |     | |
Db       1267 AAGGCAGACGAGGACAAGGCTGACTCTGATGCCGAGGATAGCAAGGACGACTCTGATGAT 1326

Qy        358 GGTAATAACGCCGCGCT 374
              |  ||  ||| || |||
Db       1327 GAGAAGCACGACGAGCT 1343


RESULT 10
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 1.5;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||     |      |||| | |    | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
              |||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 11
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674A
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 06/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674A-593

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 1355;
  Best Local Similarity   25.2%;  Pred. No. 3;
  Matches   39;  Conservative   50;  Mismatches   66;  Indels    0;  Gaps    0;

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137
              :||:||:||: ::  :  :  : ::||:||:  ::::   ||:   | :::   :   |
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197
              :  :: :  :  :||::::| :: :| :|   |:| :     |:|     ||    :
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
```

```
                : |:  : |:|| ||: |:| :| : |:|  ||:|
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYG 922
```

RESULT 12
US-10-767-701-7332/c
; Sequence 7332, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 7332
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS77572_1
US-10-767-701-7332

Query Match 6.4%; Score 29.2; DB 6; Length 531;
Best Local Similarity 48.8%; Pred. No. 2.3;
Matches 79; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```
Qy        175 CTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGC 234
              |||||   | || ||  |     | |      | || |||    || | | | |  || |
Db        475 CTGCAGGTCCATCCCGATCCGGTTCACGAACACCTTCCCCGTGCCGTTCAGGATCCCGCC 416

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 294
              |||  | ||||  | ||  |    | | ||||      |   | |||||||  | |||
Db        415 CCCGCCGGAGGCGTAGTTGACGCCGTTCAGGATCGCGCCGCCGCTCGTGTTGGGAGCCAG 356

Qy        295 GAAATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGAT 336
              |||  |   |  | || ||    || |    |   | |||||
Db        355 GAACGGCGGCGAGTAGTCCGCCTGACCCAGCATCTCGCCGAT 314
```

RESULT 13
US-10-767-471-10809/c
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
; LENGTH: 290892
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(290892)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

Query Match 6.4%; Score 29.2; DB 6; Length 290892;
Best Local Similarity 53.5%; Pred. No. 40;
Matches 61; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy        182 GCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATG 241
              | |||| | || |   ||| |    |||||   |    ||| |||||   || |  ||||
Db     234207 GTGATGACGGTGATGGTGATAGCTACATTAGTGATGTTGGTAATGGTGGTGGTGATGATG 234148

Qy        242 TAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATG 295
```

```
                  || | |     ||| || ||||| |    | ||   ||| || |    |||
Db     234147 ATGGTGATGCCAATGATTATGATGATGATGATGGTAATGGTGATGATGGTGATG 234094
```

RESULT 14
US-10-781-014-555/c
; Sequence 555, Application US/10781014
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN CARBON METABOLISM AND ENERGY
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: BGI-126CPCN
; CURRENT APPLICATION NUMBER: US/10/781,014
; CURRENT FILING DATE: 2004-02-17
; PRIOR APPLICATION NUMBER: US 09/602,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/143,208
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/151,572
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19931412.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931413.6
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931424.1
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931428.4
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 555
; LENGTH: 1682
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1659)
; OTHER INFORMATION: FRXA02209
US-10-781-014-555

Query Match 6.4%; Score 29; DB 6; Length 1682;
Best Local Similarity 63.8%; Pred. No. 4.6;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 294
              || ||||| ||| | ||||| |   || |||| |   ||||||||     ||  |||||
Db       1209 GCGGATGTTGGCGAAGGTGCCGCGCATCATGACCTCGTGGTTACCACGCCTGGAACCCAG 1150

Qy        295 GAAATGGCA 303
              | | | | |
Db       1149 GGAGTTGTA 1141
```

RESULT 15
US-10-781-014-553/c
; Sequence 553, Application US/10781014
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor

; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN CARBON METABOLISM AND ENERGY
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: BGI-126CPCN
; CURRENT APPLICATION NUMBER: US/10/781,014
; CURRENT FILING DATE: 2004-02-17
; PRIOR APPLICATION NUMBER: US 09/602,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/143,208
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/151,572
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19931412.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931413.6
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931424.1
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931428.4
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 553
; LENGTH: 1694
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1671)
; OTHER INFORMATION: RXN02209
US-10-781-014-553

Query Match 6.4%; Score 29; DB 6; Length 1694;
Best Local Similarity 63.8%; Pred. No. 4.6;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 294
              || ||||| ||| | ||||| |   || |||| |   ||||||||     ||  |||||
Db       1221 GCGGATGTTGGCGAAGGTGCCGCGCATCATGACCTCGTGGTTACCACGCCTGGAACCCAG 1162

Qy        295 GAAATGGCA 303
              | | | | |
Db       1161 GGAGTTGTA 1153
```

Search completed: March 15, 2004, 13:47:23
Job time : 56.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 02:55:53 ; Search time 1963.4 Seconds
(without alignments)
6935.497 Million cell updates/sec

Title: US-09-543-407-21
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	231	50.7	680	12	BJ618688	BJ618688 BJ618688
c	2	39.8	8.7	484	9	AV430994	AV430994 AV430994
c	3	38.4	8.4	759	28	BH202614	BH202614 Sm1-58F3.
c	4	35.2	7.7	796	28	BH180509	BH180509 016_P_06-
c	5	35.2	7.7	796	29	CNS07MCJ	AL617461 T7 end of
	6	35	7.7	497	29	CG449269	CG449269 OG7AE12TV
c	7	35	7.7	552	29	CG448618	CG448618 OG7AE11TH
c	8	34.8	7.6	527	28	BH207612	BH207612 Sm1-57B8.
	9	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	10	34	7.5	234	10	BB581987	BB581987 BB581987
c	11	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
	12	34	7.5	1201	13	BX381961	BX381961 BX381961
	13	33.8	7.4	359	10	BE942582	BE942582 EST422161
c	14	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	15	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	16	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
	17	33.8	7.4	712	9	AL873397	AL873397 AL873397
	18	33.8	7.4	935	14	CK017088	CK017088 AGENCOURT
	19	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	20	33.8	7.4	1013	29	CNS02IDX	AL198798 Tetraodon
c	21	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	22	33.8	7.4	1124	14	CK161585	CK161585 FGAS01415
c	23	33.4	7.3	505	6	AL809250	Al809250 Triticum
c	24	33.4	7.3	512	10	BE597399	BE597399 PI1_69_A1
c	25	33.4	7.3	699	12	BI959328	BI959328 HVSMEn001
	26	33.4	7.3	715	13	BU444683	BU444683 603765715
c	27	33.4	7.3	740	14	CK280834	CK280834 EST743556
	28	33.2	7.3	629	9	AU295325	AU295325 AU295325
	29	33.2	7.3	788	9	AU139321	AU139321 AU139321
	30	33.2	7.3	875	14	CD360922	CD360922 AGENCOURT
c	31	33.2	7.3	1646	10	BF579120	BF579120 602096054
c	32	33	7.2	548	28	BH200391	BH200391 Sm1-39D16
	33	33	7.2	649	12	BI261742	BI261742 602954071
	34	33	7.2	965	29	CNS02W4D	AL216598 Tetraodon
c	35	32.8	7.2	429	12	BI537361	BI537361 397161 MA
c	36	32.8	7.2	445	10	BF586449	BF586449 FM1_28_C0
	37	32.8	7.2	485	14	CA561783	CA561783 K0291H09-
c	38	32.8	7.2	521	14	CA703231	CA703231 wdk1c.pk0
	39	32.8	7.2	572	14	CF915995	CF915995 B0988D03-
	40	32.8	7.2	660	14	CF174513	CF174513 B0940C01-
	41	32.8	7.2	733	14	CB967076	CB967076 NL47_A12
	42	32.8	7.2	830	9	AJ568520	AJ568520 AJ568520
c	43	32.8	7.2	969	13	BU197195	BU197195 AGENCOURT
c	44	32.6	7.1	250	28	BH181760	BH181760 019_K_18-
c	45	32.6	7.1	250	29	CNS07NB8	AL618710 T3 end of

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 50.7%; Score 231; DB 12; Length 680;
Best Local Similarity 69.1%; Pred. No. 1.9e-53;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || |||||| || | ||||||||||||||| ||    || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||  | ||| |||||| |||||||||  || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||    ||| |||||||||    ||     ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
              || || |||||  | ||| |                | ||||   ||    |  | | |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||    |   || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
AV430994/c
LOCUS AV430994 484 bp mRNA linear EST 23-AUG-2000
DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone PL027e10_r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
REFERENCE 1 (bases 1 to 484)
AUTHORS Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and Tabata,S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
PUBMED 10907854
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PL027e10_r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 8.7%; Score 39.8; DB 9; Length 484;
Best Local Similarity 54.4%; Pred. No. 3.2;
Matches 80; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy        293 ATGAAATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAAT 352
              | | |  ||| | |||  |   |   | | ||| || |  | |  | | ||  |||
Db        469 AAGCAGAGGCTCTTGCCAGCGCCAGTTGCCCAAAACACACAAGTGTCCCGGCCGCTCAAG 410

Qy        353 ACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGC 412
              |||||  ||| ||||| ||||||  |  | ||  ||   ||  |||| ||| ||||||
Db        409 ACGGCCTCAATCACGCCTCGCTGGCCAGGCCGAAAGCTGGTGTATCCAAAGTACTGGTGC 350

Qy        413 GTCAGGTTGGTTTTGGCAACAACGCCA 439
                  ||    ||| |  || ||
Db        349 AGGGTGCGGGAAAGGGCGTCGCCGGCA 323
```

RESULT 3
BH202614/c
LOCUS BH202614 759 bp DNA linear GSS 15-NOV-2002
DEFINITION Sm1-58F3.TR Sm1 Schistosoma mansoni genomic clone Sm1-58F3, genomic survey sequence.
ACCESSION BH202614
VERSION BH202614.1 GI:16375293
KEYWORDS GSS.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 759)
AUTHORS Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and El-Sayed,N.M.
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sm1 BAC library for gene discovery and map construction
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: Sm1-58F3.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Schistosoma mansoni (Puerto Rico) Sm1 BAC library. For clone availability, please contact Dr. Najib El-Sayed at TIGR (nelsayed@tigr.org) or Dr. Phillip LoVerde at State University of New York, Buffalo, New York, USA (loverde@buffalo.edu)
Seq primer: M13 Rev
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .759
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
/strain="Puerto Rico"
/db_xref="taxon:6183"
/clone="Sm1-58F3"
/clone_lib="Sm1"
/note="Vector: pBeloBAC11; Site_1: Hin dIII; Constructed

in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBeloBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 X the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

ORIGIN

Query Match 8.4%; Score 38.4; DB 28; Length 759;
Best Local Similarity 47.5%; Pred. No. 9.6;
Matches 114; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | | | ||||   ||||     |    ||||  | |    |||  |   |   | |
Db        342 GATGATGCTGCCGCTGCTTATGATGATGATGATCCTGCTGCTGCTAATGCTGCTGCTGCT 283

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTT 276
                | | ||| | | |  | ||||  |   | | ||  ||| || |||||  |    | |
Db        282 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 223

Qy        277 ACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGAT 336
                || || ||   |||| ||| | | ||  |  |||| | |||| |||  |
Db        222 GATGATGATGATAAGGATGATGTGGATGAAGTTGAAGAAAGGAATGATAAAGACTTTGGA 163

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGAT 396
              |  |||       |||  |   || | |  ||    | | |  | |  | || ||||
Db        162 CATGATGTGTTTTGTGACGATGACGATGATGTAGCTGATGATGAAGATAAGGGATATGAT 103
```

RESULT 4
BH180509/c
LOCUS BH180509 796 bp DNA linear GSS 19-OCT-2001
DEFINITION 016_P_06-21 SmBAC1 Schistosoma mansoni genomic clone 016P06 5', genomic survey sequence.
ACCESSION BH180509
VERSION BH180509.1 GI:16282522
KEYWORDS GSS.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 796)
AUTHORS Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
TITLE Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library
JOURNAL Genomics 65 (2), 87-94 (2000)
MEDLINE 20247247
PUBMED 10783255
COMMENT Other_GSSs: 016_P_06-rev
Contact: Pierce RJ
INSERM U 167
Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr
CNS sequencing ID=DG0AA016DH03CP1
Plate: 016 row: P column: 06
Seq primer: M13 -21 primer
Class: BAC ends
High quality sequence stop: 796.
FEATURES Location/Qualifiers
source 1. .796
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="016P06"
/sex="mixed"
/dev_stage="cercariae"
/lab_host="Biomphalaria glabrata"
/clone_lib="SmBAC1"
/note="Vector: pBeloBAC 11; Site_1: Hind III; Partially Hind III digested and size-selected S. mansoni cercarial DNA was ligated into Hind III digested pBeloBAC 11 vector and used to transform E. coli DH10B. The complete library contains 23808 clones from 4 independent sizing-ligation-transformations. Average insert size ranges from 70-127 kb and genome coverage is 7.9-fold."

ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 796;
Best Local Similarity 46.1%; Pred. No. 77;
Matches 118; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
Qy        141 TTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGA 200
              |||| |  | | |    ||   |||||   ||||          | ||| | | |   |||
Db        353 TTATGATGATGATGATCCTGCTGCTGCTAATGCTGCTGCTGCTGCTGCTCCTGATGATGA 294

Qy        201 AACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAA 260
                      |    | | ||| | | |  |  ||||  |   | | || |||  |
Db        293 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 234

Qy        261 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAA 320
              | ||||| |   | |    || || ||   ||||  |||   | |  ||  |  || |
Db        233 TGATGATGATGATGATGATGATGATGATAAGGATGATGTGGATGAAGTTGAAGAAAGGGA 174

Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | ||||  | |  |  | |  |||         |||  |   || | |  ||    | | |
Db        173 TGATAAAGATTTTGGAAATGATGTGTTTTGTGACGATGACGATGATGTAGCTGATGATGA 114

Qy        381 TCAGACCGCATCTGAT 396
              | |  | ||  ||||
Db        113 AGATAAGGGATATGAT 98
```

RESULT 5
CNS07MCJ/c
LOCUS CNS07MCJ 796 bp DNA linear GSS 30-NOV-2001
DEFINITION T7 end of clone 016DH03 of library SmBAC1 from strain Puerto-Rican of Schistosoma mansoni, genomic survey sequence.
ACCESSION AL617461
VERSION AL617461.1 GI:16030919
KEYWORDS GSS.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 796)
AUTHORS Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
TITLE Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library
JOURNAL Genomics 65 (2), 87-94 (2000)
MEDLINE 20247247
PUBMED 10783255
REFERENCE 2 (bases 1 to 796)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Partially Hind III digested and size-selected S. mansoni cercarial DNA was ligated into Hind III digested pBeloBAC 11 vector and used to transform E. coli DH10B. The complete library contains 23808 clones from 4 independent sizing-ligation-transformations. Average insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
FEATURES Location/Qualifiers
source 1. .796
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"

/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="016DH03"
/clone_lib="SmBAC1"
/note="end : T7"

ORIGIN

Query Match 7.7%; Score 35.2; DB 29; Length 796;
Best Local Similarity 46.1%; Pred. No. 77;
Matches 118; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
Qy      141 TTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGA 200
Db      353 TTATGATGATGATGATCCTGCTGCTGCTAATGCTGCTGCTGCTGCTGCTCCTGATGATGA 294

Qy      201 AACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAA 260
Db      293 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 234

Qy      261 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAA 320
Db      233 TGATGATGATGATGATGATGATGATGATAAGGATGATGTGGATGAAGTTGAAGAAAGGGA 174

Qy      321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
Db      173 TGATAAAGATTTTGGAAATGATGTGTTTTGTGACGATGACGATGATGTAGCTGATGATGA 114

Qy      381 TCAGACCGCATCTGAT 396
Db      113 AGATAAGGGATATGAT 98
```

RESULT 6
CG449269
LOCUS CG449269 497 bp DNA linear GSS 17-SEP-2003
DEFINITION OG7AE12TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0843A23, genomic survey sequence.
ACCESSION CG449269
VERSION CG449269.1 GI:34834269
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 497)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .497
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0843A23"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

Query Match 7.7%; Score 35; DB 29; Length 497;
Best Local Similarity 45.9%; Pred. No. 71;
Matches 119; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy      179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
Db       18 AAAGGGAAAACCCATTTTCATATCGGGCCAGAAACGAGAACCGGTCTGGAAACCGCGGTA 77

Qy      239 ATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAA 298
Db       78 AACCCTGATCTCAAGCGAAACCATAGGATAACCTGGCACACGGTGGGTTGCTTTTCGAAA 137

Qy      299 TGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
Db      138 ATCCCAATAGAGACAGGCTGCGTGCCAACAACACTGAGCCCATGGTGGTGGTGCACGTGC 197

Qy      359 GTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGG 418
Db      198 GCATCACAGTCGTCCGCGCGTGAAAATCTGTCGCTGATACCCCCGATCGGATGGCCCAGT 257

Qy      419 TTGGTTTTGGCAACAACGC 437
Db      258 CTCGTTCTCGCGCGACCCC 276
```

RESULT 7
CG448618/c
LOCUS CG448618 552 bp DNA linear GSS 17-SEP-2003
DEFINITION OG7AE11TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0843A21, genomic survey sequence.
ACCESSION CG448618
VERSION CG448618.1 GI:34833618
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 552)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .552
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0843A21"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

Query Match 7.7%; Score 35; DB 29; Length 552;
Best Local Similarity 45.9%; Pred. No. 75;
Matches 119; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy      179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
Db      480 AAAGGGAAAACCCATTTTCAAATCGGGCCAGAAACGAGAACCGGTCTGGAAACCGCGGTA 421

Qy      239 ATGTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAA 298
Db      420 AACCCTGATCTCAAGCGAAACCATAGGATAACCTGGCACACGGTGGGTTGCTTTTCGAAA 361
```

```
Qy        299 TGGCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
              |  ||  ||||  |  |   || || ||| | ||    |  || |       |||
Db        360 ATCCCAATAGAGACAGGCTGCGTGCCAACAACACTGAGCCCATGGTGGTGGTGCACGTGC 301

Qy        359 GTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGG 418
              | |  |  | ||  |  |      |  | |   ||||| ||  ||    | ||   |||
Db        300 GCATCACAGTCGTCCGCGCGTGAAAATCTGTCGCTGATACCCCCGATCGGATGGCCCAGT 241

Qy        419 TTGGTTTTGGCAACAACGC 437
              | ||| | ||   | | |
Db        240 CTCGTTCTCGCGCGACCCC 222
```

RESULT 8
BH207612/c
LOCUS BH207612 527 bp DNA linear GSS 15-NOV-2002
DEFINITION Sm1-57B8.TF Sm1 Schistosoma mansoni genomic clone Sm1-57B8, genomic survey sequence.
ACCESSION BH207612
VERSION BH207612.1 GI:16384748
KEYWORDS GSS.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 527)
AUTHORS Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and El-Sayed,N.M.
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sm1 BAC library for gene discovery and map construction
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: Sm1-57B8.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Schistosoma mansoni (Puerto Rico) Sm1 BAC library. For clone availability, please contact Dr. Najib El-Sayed at TIGR (nelsayed@tigr.org) or Dr. Phillip LoVerde at State University of New York, Buffalo, New York, USA (loverde@buffalo.edu)
Seq primer: M13 For
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .527
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
/strain="Puerto Rico"
/db_xref="taxon:6183"
/clone="Sm1-57B8"
/clone_lib="Sm1"
/note="Vector: pBeloBAC11; Site_1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBeloBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 X the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."
ORIGIN

Query Match 7.6%; Score 34.8; DB 28; Length 527;
Best Local Similarity 46.3%; Pred. No. 83;
Matches 114; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

```
Qy        151 GGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATT 210
              | | | ||| | | ||    ||||         || |||   | |   |||      |  |
Db        348 GCTGCTGCTGATGATGATGATGCTGATGCTGCCGCTGCTTATGATGATGATGATCCTGCT 289

Qy        211 ACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATTATGATCAG 270
              |        | | ||| | | |   ||||  |   | | ||  ||| || ||||| |
Db        288 GCTGCTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 229

Qy        271 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGACCAGTGGAACGCTAAAAAC 330
              | |      || || ||   ||||  |||    | |   ||  |  || | | |||| |
Db        228 GATGATGATGATGATGATAAGGATGATGTGGATGAAGTTGAAGAAAGGGATGATAAAGAT 169

Qy        331 TCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCA 390
              | |  | |  |||        |||  |   || | |  ||    | | |  | |  | |
Db        168 TTTGGAAATGATGTGTTTTGTGACGATGACGATGATGTAGCTGATGATGAAGATAAGGGA 109

Qy        391 TCTGAT 396
              | ||||
Db        108 TATGAT 103
```

RESULT 9
BG671547
LOCUS BG671547 1036 bp mRNA linear EST 30-APR-2001
DEFINITION DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5', mRNA sequence.
ACCESSION BG671547
VERSION BG671547.1 GI:13893646
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1036)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
TITLE Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
MEDLINE 22056133
PUBMED 12060780
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR PRimers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1036
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRNBUB05"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/clone_lib="Rat DRG Library"
ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 1036;
Best Local Similarity 49.2%; Pred. No. 1.6e+02;

```
  Matches   90;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy         11 TAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTC 70
Db        168 TGAAAGATACATGAGTGACAGTGTCCGTGATTGCAGACATTGCTGTATGATGCGTCCTGT 227

Qy         71 CACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAA 130
Db        228 CTGTAGGAGTCCTCGCTGGTAACGATGCCGAGAGTGCACATCGATATCGGTGGGCGGTTA 287

Qy        131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
Db        288 CATCTGAAGTGGAAGAGGTGGCTCTTTCGAACTCTGACTTTGATCTGTCAAGGGACGCTT 347

Qy        191 GTA 193
Db        348 GTA 350


RESULT 10
BB581987/c
LOCUS       BB581987                 234 bp    mRNA    linear   EST 30-NOV-2000
DEFINITION  BB581987 RIKEN full-length enriched, adult male colon Mus musculus
            cDNA clone 9030005005 5', mRNA sequence.
ACCESSION   BB581987
VERSION     BB581987.1  GI:11478531
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
            Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
            Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
            Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
            Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
            Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9030005005"
                     /sex="male"
                     /tissue_type="colon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male colon"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 10;  Length 234;
  Best Local Similarity   51.3%;  Pred. No. 99;
  Matches   79;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         35 TCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATC 94
Db        200 TAGTTTTTGGTAGGTCGGTTTATGCGGGTTTCTTCTTCCACCAGGCCCGCCGCCGTTTCC 141

Qy         95 ATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTT 154
Db        140 CGAACGCCGGGTGCTTTATAGCCTTGTTCAAAATATGGTAGCGCATTTACCACCACCTTT 81

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
Db         80 CCGGTTACCAAGCCCGTGCCCGCCATTGCGCGGC 47


RESULT 11
CA711477/c
LOCUS       CA711477                 500 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wdk2c.pk014.o3 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.o3 5'
            end, mRNA sequence.
ACCESSION   CA711477
VERSION     CA711477.1  GI:25433270
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Triticum aestivum"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone="wdk2c.pk014.o3"
                     /tissue_type="kernel"
                     /clone_lib="wdk2c"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
                     days after anthesis."
ORIGIN

  Query Match             7.5%;  Score 34;  DB 14;  Length 500;
  Best Local Similarity   54.5%;  Pred. No. 1.4e+02;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||   |  ||| ||||||||   ||   |  | ||| ||       | | ||  | ||| |
Db        274 AGNTGNTGCAGTCGCAGCCNCCGGTCTCCCCGGCCGTNTCGAACCCGCCGGAGCTCCCCN 215

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGAC 126
                 ||| ||||| ||  || |  | |||||||||     |    ||||   ||
Db        214 TGTGGGAGGCGGTGGGCATGACCATGGCGGCGGCGCCGGCGGTGGCCNCCAC 163


RESULT 12
BX381961
LOCUS       BX381961                1201 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI072YF05 3-PRIME, mRNA sequence.
ACCESSION   BX381961
VERSION     BX381961.1  GI:30453007
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI072CC03NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI072YF05"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             7.5%;  Score 34;  DB 13;  Length 1201;
  Best Local Similarity   6.9%;  Pred. No. 2e+02;
  Matches   30;  Conservative  190;  Mismatches  217;  Indels    0;  Gaps    0;

Qy          2 TGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTG 61
              |   ::  :             | ::      :  :      :      :          :
Db        725 TNNKMMNNKNNNNNNMMNNCNCNKKKNNKTNMNNMMNNNNNKKNNNNNKNNNNNNNNNNK 784

Qy         62 GCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCC 121
              ::: :: :   :  :    |:: :::::  : |:  :: : ::|  :   :  ::| ::|
Db        785 KMKMSKCKKKKMCCKMCCMCKKKKKMBKGKMVCMCKMMMKNKKCMCMKBMMMCKMCMKMC 844

Qy        122 CGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAA 181
              :::   :::: : :: :: :   :::::     ::   : : :::  : ::  |    :
Db        845 MBKKNMMMMMMKMKMMMNKNKAMMKKDNMMMMMMCAMMCMMMKKNMMMBKKNCMMKKMC 904

Qy        182 GCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATG 241
              :  :: : : :: ::     : :|   :::  : :: ::  : :: ::  ::::|  ::
Db        905 CMMKKKKHKKNMMKKMMMCKKBMMCKKNKMMNKMMVKKKKAMMNBKCVMKMKMKMCCKKV 964

Qy        242 TAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGG 301
              ::::||:::: |  :: :  : : ::  |::::   :| ::     :::|| :::: :
Db        965 MMBKMCAKKKKMCMCVKKVMCMCDBKMCMCKKKBMKKVCCKKMCMMMMMMCAMKVMMMBA 1024

Qy        302 CACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
              ::::: :::::::    :::  ::::   ::|: :          :::| :: :  ::|
Db       1025 MMMMBAMVVMMMMNNMMMMMCKKMVMKKKKMCKKKBMSKKMMMNKKVCKMBBMCMSKGGC 1084

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTG 421
              :   | | ||:      : : :  |   ::::   ||: : : : ::::|::  :: ::
Db       1085 MMCGCMCKGCVGCMNSCCMMBGSKGCCKKKMBKCMMCCMCMCBCMCBKKVCVKKKKKCKK 1144

Qy        422 GTTTTGGCAACAACGCC 438
              ::::::    | : ::|
Db       1145 KKKKKKCKKCCNMKKBC 1161


RESULT 13
BE942582
LOCUS       BE942582                359 bp    mRNA    linear   EST 03-OCT-2000
DEFINITION  EST422161 MGHG Medicago truncatula cDNA clone pMGHG-8L14, mRNA
            sequence.
ACCESSION   BE942582
VERSION     BE942582.1  GI:10520341
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Cote,F., Ojanen-Reuths,T., Hahn,M.G., VandenBosch,K., Hur,J.,
            Beremand,P., Endre,G., Town,C.D., Bowman,C.L., Craven,M.B. and
            Cho,J., Fraser,C.M.
  TITLE     ESTs from seedling roots of Medicago truncatula after treatment
            with beta glucan elicitor preparation from Phytophthora sojae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michael G. Hahn
            Complex Carbohydrate Research Center
            University of Georgia
            220 Riverbend Road, Athens, GA 30602-4712, USA
            Tel: 706-542-4457
            Fax: 706-542-4412
            Email: hahn@ccrc.uga.edu
            University of Georgia name: G270434e TIGR sequence name: MTJAX67TK
            More information is available at: http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .359
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pMGHG-8L14"
                     /tissue_type="Roots from four day old seedlings"
                     /dev_stage="2 days after treatment with beta glucan
                     elicitor preparation from Phytophthora sojae"
                     /lab_host="E. coli strain XLOLR"
                     /clone_lib="MGHG"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged  using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
```

from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

ORIGIN

```
  Query Match             7.4%;  Score 33.8;  DB 10;  Length 359;
  Best Local Similarity   51.3%;  Pred. No. 1.3e+02;
  Matches   77;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        248 AGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 307
              ||| ||     |||||||||||| |  | | | | |||| ||| |    |||| | ||  ||
Db          1 AGGTTGTNACAAATTATGATCAACAAGATTCACATGTTATTAGCTCAGAAAGGTCAATTG 60

Qy        308 CAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACG 367
              | ||     |  | || |||  ||| | ||||           | | || | ||
Db         61 TGAAACAAGAATATCCCAACAACGACGACAGTACTACTCCGAGGAGTGATGTTACTGACA 120

Qy        368 CCGCGCTGGTTAATCAGACCGCATCTGATT 397
              |   ||| | | |      |||||||
Db        121 ACTTATTGGATCCTAATTTGTGGTCTGATT 150
```

RESULT 14
AA746477/c

```
LOCUS       AA746477                365 bp    mRNA    linear   EST 27-JAN-1998
DEFINITION  nw62e02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
            similar to TR:O00555 O00555 NEURONAL CALCIUM CHANNEL ALPHA 1A
            SUBUNIT ISOFORM A-1. ;contains element TAR1 repetitive element ;,
            mRNA sequence.
ACCESSION   AA746477
VERSION     AA746477.1  GI:2786463
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 661    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 292.
FEATURES                Location/Qualifiers
     source             1. .365
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1251194"
                        /tissue_type="germinal center B cell"
                        /lab_host="DH10B"
                        /clone_lib="NCI_CGAP_GCB1"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                        was prepared from human tonsillar cells enriched for
                        germinal center B cells by flow sorting (CD20+, IgD-),
                        provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                        (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                        primed with a Not I - oligo(dT) primer
                        [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                        ]. Double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7T3 vector. Library
                        went through one round of normalization, and was
                        constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 365;
  Best Local Similarity   56.9%;  Pred. No. 1.4e+02;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  ||||||   ||||||   | ||   | ||| |||  | ||| |||    |  ||||
Db        119 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGTGGCCAGGCCGGCCGGGCGCCACC 60

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |  || |  ||| ||||  ||   | ||   | | |   | | |||| |
Db         59 AGCGGCCCTCGGAGGTACCCAGGCCCGCACGGCCGACGCTACTGGCCGG 11
```

RESULT 15
CD374421/c

```
LOCUS       CD374421                558 bp    mRNA    linear   EST 19-AUG-2003
DEFINITION  TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)
            Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION   CD374421
VERSION     CD374421.1  GI:31229939
KEYWORDS    EST.
SOURCE      Wuchereria bancrofti
  ORGANISM  Wuchereria bancrofti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Wuchereria.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
  TITLE     Toward the characterization of the Wuchereria bancrofti
            microfilarial transcriptome with comparisons to those of Brugia
            malayi and Onchocerca volvulus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Thomas B. Nutman
            Laboratory of Parasitic Diseases
            NIH
            Building 4 Room 126, Bethesda, MD 20892-0425, USA
            Email: tnutman@niaid.nih.gov
            Seq primer: pBluescript T3.
FEATURES                Location/Qualifiers
     source             1. .558
                        /organism="Wuchereria bancrofti"
                        /mol_type="mRNA"
                        /db_xref="taxon:6293"
                        /clone="TNWbmfC2H3"
                        /dev_stage="microfilaria"
                        /lab_host="XL1-Blue MRF'"
                        /clone_lib="Wuchereria bancrofti microfilaria cDNA
                        (SAW95SjL-WbMf)"
                        /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                        Xho I; Lymphatic filarial nematode parasite of humans.
                        mRNA was prepared from approximately 85,000 microfilariae
                        isolated from the blood of an infected individual from
                        Guyana, South America and converted to double-stranded
                        cDNA using reverse transcriptase and oligo(dT) followed by
                        RNase H and DNA pol I. The library has 2.9 x 10E5
                        independent recombinants and the average insert size is
                        ~1kb. The library was constructed by Sandra J. Laney. The
                        library is available from Dr.S.A.Williams, email:
                        genome@smith.edu."

ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 558;
  Best Local Similarity   60.7%;  Pred. No. 1.6e+02;
  Matches   65;  Conservative    0;  Mismatches   40;  Indels    2;  Gaps    1;

Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              ||| |||| |||||||| |     || ||||| |||| |    | ||  | |
Db        161 TGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCCCAAATTCTANGCT 102
```

```
Qy        136 AGCATTTATC--AGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||  |  ||  ||||  ||  | | ||||  |||||||||  ||
Db        101 GACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55
```

Search completed: March 15, 2004, 10:09:20
Job time : 1969.4 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 19:02:18 ; Search time 2269.8 Seconds
(without alignments)
8707.562 Million cell updates/sec

Title: US-09-543-407-23
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		404.8	88.8	456	6	I44909	I44909 Sequence 58
2		404.8	88.8	2067	1	SEU43280	U43280 Salmonella
3		396.8	87.0	5103	1	STAJ2301	AJ002301 Salmonell
4		396.8	87.0	22411	1	AE008749	AE008749 Salmonell
5		395.2	86.7	254050	1	AL627269	AL627269 Salmonell
6	c	395.2	86.7	301983	1	AE016840	AE016840 Salmonell
7		389.4	85.4	1048	1	STAGFBA	AJ000514 Salmonell
8		283.2	62.1	2889	1	CSP515700	AJ515700 Citrobact
9		280.6	61.5	361	6	I44908	I44908 Sequence 56
10		243.6	53.4	2920	1	CFR515701	AJ515701 Citrobact
11		240	52.6	4680	1	ECCSGABDG	X90754 E.coli csgG
12		240	52.6	10346	1	AE000205	AE000205 Escherich
13		240	52.6	15047	1	D90741	D90741 Escherichia
14		238.4	52.3	456	6	AX814811	AX814811 Sequence
15		238.4	52.3	648	1	ECOCSGAA	L04979 Escherichia
16		233.4	51.2	306358	1	AE016759	AE016759 Escherich
17		231.8	50.8	1711	1	AF275733	AF275733 Escherich
18		231.8	50.8	10190	1	AE005315	AE005315 Escherich
19		231.8	50.8	327773	1	AP002554	AP002554 Escherich
20		229	50.2	230	1	SEU53207	U53207 Salmonella
21		193.6	42.5	2883	1	ESA515702	AJ515702 Enterobac
22		171.2	37.5	10370	1	AE015131	AE015131 Shigella
23		171.2	37.5	292504	1	AE016981	AE016981 Shigella
24	c	148.2	32.5	437	1	AF237726	AF237726 Shigella
25		77.2	16.9	19201	1	D90742	D90742 Escherichia
26		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27		45.8	10.0	78	6	AX814809	AX814809 Sequence
28		44.8	9.8	72	6	AX814798	AX814798 Sequence
29	c	40	8.8	2000	6	AX655393	AX655393 Sequence
30	c	37.6	8.2	110000	10	AE014175_0	AE014175 Mus muscu
31	c	37.6	8.2	156823	2	AC102815	AC102815 Mus muscu
32		37	8.1	306	8	PCU34351	U34351 Pneumocysti
33	c	37	8.1	2196	6	E03183	E03183 cDNA sequen
34	c	37	8.1	2199	9	AF141383	AF141383 Homo sapi
35	c	37	8.1	2374	9	HUMAARE	D38441 Human mRNA
36	c	37	8.1	2775	9	BC001499	BC001499 Homo sapi
37	c	37	8.1	2775	9	BC001826	BC001826 Homo sapi
38	c	37	8.1	2798	9	BC000362	BC000362 Homo sapi
39	c	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
40	c	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
41	c	36	7.9	168267	2	AC146009	AC146009 Pan trogl
42		36	7.9	226967	2	AC113654	AC113654 Rattus no
43		35.6	7.8	301214	1	AE016786	AE016786 Pseudomon
44	c	35.4	7.8	76000	9	AC092793	AC092793 Homo sapi
45	c	35.4	7.8	155001	2	AC013827	AC013827 Homo sapi

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 3.6e-109;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match           88.8%;  Score 404.8;  DB 1;  Length 2067;
  Best Local Similarity   93.0%;  Pred. No. 4e-109;
```

```
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||      ||   ||     |
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               || |    || |  |  ||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match             87.0%;  Score 396.8;  DB 1;  Length 5103;
  Best Local Similarity   91.9%;  Pred. No. 1e-106;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||| |||     ||    |     |
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES Location/Qualifiers

source 1. .22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

gene 434. .1308
/gene="phoH"
/note="synonym: STM1126"

RBS 434. .439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"

CDS 454. .1308
/gene="phoH"
/note="similar to E. coli PhoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=11
/product="PhoB-dependent ATP-binding pho regulon component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
QRTLHAYS"

gene complement(1414. .2303)
/gene="STM1127"

CDS complement(1414. .2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29% identity in aa 15 - 264"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
QARESKLRTMNALDMTK"

RBS complement(2298. .2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"

gene complement(2580. .4088)
/gene="STM1128"

CDS complement(2580. .4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22% identity in aa 7 - 478"
/codon_start=1
/transl_table=11
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"

RBS complement(4083. .4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"

gene complement(4413. .5099)
/gene="STM1129"

CDS complement(4413. .5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 - 226"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
WFSHAVKR"

RBS complement(5094. .5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"

gene 5599. .6759
/gene="STM1130"

CDS 5599. .6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41% identity in aa 40 - 403"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"

gene 6793. .7497
/gene="STM1131"

RBS 6793. .6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           87.0%;  Score 396.8;  DB 1;  Length 22411;
  Best Local Similarity   91.9%;  Pred. No. 1.2e-106;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||| |||     ||    |    |
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
  PUBMED    11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

/note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature"
gene 3001. .4011
/gene="STY1079"
/note="synonym: pyrD"
CDS 3001. .4011
/gene="STY1079"
/note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q8Z7S9"
/db_xref="SWISS-PROT:Q8Z7S9"
/translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
HI"
misc_feature 3136. .3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHOdehase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
misc_feature 3238. .3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
misc_feature 3874. .3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
gene 4174. .4719
/gene="STY1080"
CDS 4174. .4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepency in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q8Z7S8"
/translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
IMNDRLKPQVHCHSFSLEQAV"
gene complement(4716. .5825)
/gene="STY1081"
CDS complement(4716. .5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Fasta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q8Z7S7"
/db_xref="SPTREMBL:Q8Z7S7"
/translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
misc_feature complement(4737. .4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
gene 5924. .8032
/gene="STY1082"
CDS 5924. .8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q8Z7S6"
/db_xref="SPTREMBL:Q8Z7S6"
/translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
LSPDFARNRQIHNCWLIRAA"
misc_feature 5927. .7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, score 781.30, E-value 3.7e-231"
misc_feature 6506. .6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
gene 8045. .9952
/gene="STY1083"
CDS 8045. .9952
/gene="STY1083"
/note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa overlap"

Query Match 86.7%; Score 395.2; DB 1; Length 254050;
Best Local Similarity 91.7%; Pred. No. 4.2e-106;
Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```
Qy        1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy       61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy      121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
            ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy      181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy      241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
            |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db    89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||||||||||||||||||||||||||||||| |||       ||    |     |
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361
```

```
RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.7%;  Score 395.2;  DB 1;  Length 301983;
Best Local Similarity   91.7%;  Pred. No. 4.3e-106;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||| |||   ||    |   |
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               || |    || |  | ||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match           85.4%;  Score 389.4;  DB 1;  Length 1048;
  Best Local Similarity  91.0%;  Pred. No. 1.5e-104;
  Matches  414;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||| |||      ||     |     |
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047
```

```
RESULT 8
CSP515700
LOCUS       CSP515700               2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           62.1%;  Score 283.2;  DB 1;  Length 2889;
  Best Local Similarity  77.9%;  Pred. No. 6.6e-73;
  Matches  355;  Conservative    0;  Mismatches   98;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| |||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| ||||||||||||||||||   |     ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  ||||||||||||||  ||   |  |||||| ||||||||  ||||||
```

```
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571


RESULT 9
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;
FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           61.5%;  Score 280.6;  DB 6;  Length 361;
  Best Local Similarity  89.9%;  Pred. No. 3.4e-72;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123

Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183

Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243

Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303

Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 363

Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 GAAATGGCACATGCAAATCAGACCGCATCTGATTC 398

Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 10
CFR515701
LOCUS       CFR515701                2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
  Query Match          53.4%;  Score 243.6;  DB 1;  Length 2920;
  Best Local Similarity  72.6%;  Pred. No. 4.1e-61;
  Matches  331;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||| | |||| || || || || || ||     ||| ||||   ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || |||||| |||||||| ||    ||||| || ||  |||  ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||| ||| | ||||| |||   |||| ||||  || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||||| |||||  |||||||| || |||||   ||||      || ||||   | | | |
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |   |    ||    | |||||||| || ||||| ||||  ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || |||||||| | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

```
RESULT 11
ECCSGABDG
LOCUS       ECCSGABDG                 4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
```

```
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           52.6%;  Score 240;  DB 1;  Length 4680;
  Best Local Similarity  70.4%;  Pred. No. 5e-60;
  Matches  321;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| ||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || ||  | || |||||||||| ||||| | |||   |  |||||  |   | | |
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    |  | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 12
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
```

```
          and products; all new functional assignments courtesy of Monica
          Riley; added promoters, protein binding sites, and repeated
          sequences described in reference 1. The unique numeric identifiers
          beginning with a lowercase 'b' assigned to each gene (protein- or
          RNA-encoding) are now designated as gene synonyms instead of
          labels. This should allow them to be searched for in Entrez as gene
          names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

Query Match              52.6%;  Score 240;  DB 1;  Length 10346;
Best Local Similarity    70.4%;  Pred. No. 5.4e-60;
Matches  321;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||    ||| |||||||||    ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || ||  | || |||||||||  ||||| | |||   |  |||||  |    | | |
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |    ||    |   | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

RESULT 13
D90741
LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
FEATURES Location/Qualifiers
source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER

```
                ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS             2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
CDS             3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
gene            complement(4117. .4950)
                /gene="csgG"
CDS             complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
gene            complement(4977. .5393)
                /gene="csgF"
CDS             complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene            complement(5418. .5807)
                /gene="csgE"
CDS             complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
gene            complement(5812. .6462)
                /gene="csgD"
CDS             complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene            7217. .7672
                /gene="csgB"
CDS             7217. .7672
```

Query Match 52.6%; Score 240; DB 1; Length 15047;
Best Local Similarity 70.4%; Pred. No. 5.5e-60;
Matches 321; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | ||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||||| |||||   ||| |||||||||    ||     ||||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || ||  | || ||||||||||  ||||| | |||  |  |||||  |  | | |
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |    ||    |  | ||||| |||||| | |||  ||| |  |||  ||||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168
```

```
RESULT 14
AX814811
LOCUS       AX814811                456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
```

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           52.3%;  Score 238.4;  DB 6;  Length 456;
  Best Local Similarity  70.2%;  Pred. No. 1.3e-59;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| ||||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || ||  | || ||||||||||  ||||| | |||   |  |||||  |   | | |
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |    ||     |  | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                 648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match           52.3%;  Score 238.4;  DB 1;  Length 648;
  Best Local Similarity  70.2%;  Pred. No. 1.3e-59;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| ||||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || ||  | || ||||||||||  ||||| | |||   |  |||||  |   | | |
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |    ||     |  | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 15, 2004, 07:24:07
Job time : 2273.2 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 18:52:53 ; Search time 290.4 Seconds
(without alignments)
6670.724 Million cell updates/sec

Title: US-09-543-407-23
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64628	Aac64628 AgfA::PT3
	2	404.8	88.8	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	404.8	88.8	456	2	AAT74142	Aat74142 Salmonell
	4	404.8	88.8	456	3	AAC64617	Aac64617 Salmonell
	5	382.4	83.9	456	3	AAC64623	Aac64623 AgfA::PT3
	6	377.6	82.8	456	3	AAC64631	Aac64631 AgfA::PT3
	7	361.6	79.3	456	3	AAC64626	Aac64626 AgfA::PT3
	8	355.2	77.9	456	3	AAC64625	Aac64625 AgfA::PT3
	9	353.6	77.5	456	3	AAC64622	Aac64622 AgfA::PT3
	10	353.6	77.5	456	3	AAC64629	Aac64629 AgfA::PT3
	11	352	77.2	456	3	AAC64624	Aac64624 AgfA::PT3
	12	350.4	76.8	456	3	AAC64630	Aac64630 AgfA::PT3
	13	350.4	76.8	456	3	AAC64627	Aac64627 AgfA::PT3
	14	280.6	61.5	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	280.6	61.5	361	2	AAT74141	Aat74141 Salmonell
	16	238.4	52.3	456	3	AAC64619	Aac64619 Escherich
	17	238.4	52.3	456	9	ACF36153	Acf36153 E. coli C
	18	226	49.6	646	2	AAQ62647	Aaq62647 Fibronect
	19	164	36.0	369	2	AAQ62646	Aaq62646 FNB curli
c	20	78	17.1	78	3	AAC64609	Aac64609 AgfA (SEF
	21	78	17.1	78	3	AAC64610	Aac64610 AgfA (SEF
	22	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	23	45.8	10.0	78	9	ACF36151	Acf36151 E. coli C
	24	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
	25	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	26	41	9.0	100	7	ACD68807	Acd68807 E. coli K
	27	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
	28	40.6	8.9	78	3	AAC64606	Aac64606 SefA (SEF
c	29	40.6	8.9	78	3	AAC64605	Aac64605 SefA (SEF
c	30	40	8.8	2000	7	ADA71938	Ada71938 Rice gene
c	31	37	8.1	2196	2	AAQ20001	Aaq20001 Encodes H
c	32	37	8.1	2374	7	ACC49472	Acc49472 Human N-a
c	33	37	8.1	2374	9	ADE31768	Ade31768 Human 658
c	34	37	8.1	2383	7	ABX63521	Abx63521 Human cDN
c	35	36.2	7.9	3411	5	AAS88526	Aas88526 DNA encod
c	36	36.2	7.9	3412	5	AAS89144	Aas89144 DNA encod
	37	35	7.7	100	7	ACD68808	Acd68808 E. coli K
	38	35	7.7	2000	7	ADA71938	Ada71938 Rice gene
	39	34	7.5	456	3	AAC64618	Aac64618 Salmonell
c	40	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	41	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	42	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	43	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	44	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	45	34	7.5	1344	6	AAD27016	Aad27016 E. amylov

ALIGNMENTS

RESULT 1
AAC64628
ID AAC64628 standard; DNA; 456 BP.
XX
AC AAC64628;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36352.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

```
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-138;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag=  a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 2;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 2.3e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||  |||    ||  ||   |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || | | ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
AAT74142
```

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match            88.8%;  Score 404.8;  DB 2;  Length 456;
  Best Local Similarity  93.0%;  Pred. No. 2.3e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||| ||| || || |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |  || | | ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
Best Local Similarity 93.0%; Pred. No. 2.3e-121;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||  |||   ||  ||  ||    |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||  |   ||  |  |  ||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 83.9%; Score 382.4; DB 3; Length 456;
Best Local Similarity 89.9%; Pred. No. 4.7e-114;
Matches 410; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||  |||   ||  ||  ||  |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |  |  |     ||  |  ||    |  || | ||||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 6
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match           82.8%;  Score 377.6;  DB 3;  Length 456;
  Best Local Similarity 89.3%;  Pred. No. 1.7e-112;
  Matches  407;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||||||  || |||  |   |         |   ||     |  | | ||  |
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               ||  |    ||  |  | ||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match             79.3%;  Score 361.6;  DB 3;  Length 456;
  Best Local Similarity   87.1%;  Pred. No. 2.8e-107;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| ||| |  ||||||  | |  | | ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||  |||  ||  ||  |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |  . || |  | ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 3.4e-105;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
```

```
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             77.5%;  Score 353.6;  DB 3;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 1.1e-104;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||| |           | | || | ||  |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 1.1e-104;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||   ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||      ||  ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                || |    || | |  ||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 3.8e-104;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |  ||  ||  | | | |  |  | |  |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 1.2e-103;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |||   |       | |  ||   |    |       |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match             76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 1.2e-103;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||| | | | || | || | | | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||      |||||||||||||||||||| ||| |  ||  ||  |
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |  || |  | ||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             61.5%;  Score 280.6;  DB 2;  Length 361;
  Best Local Similarity   89.9%;  Pred. No. 6.3e-81;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 363
              |||||||||||||||||||||||||||||  |||       ||   ||      |      ||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 GAAATGGCACATGCAAATCAGACCGCATCTGATTC 398
                |    || |   |  ||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
PT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
PT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             61.5%;  Score 280.6;  DB 2;  Length 361;
  Best Local Similarity   89.9%;  Pred. No. 6.3e-81;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 363
              |||||||||||||||||||||||||||||  |||       ||   ||      |      ||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 GAAATGGCACATGCAAATCAGACCGCATCTGATTC 398
                |    || |   |  ||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 15, 2004, 04:14:41
```

Job time : 291.4 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:01:03 ; Search time 56.8 Seconds
(without alignments)
4455.240 Million cell updates/sec

Title: US-09-543-407-23
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	404.8	88.8	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	280.6	61.5	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	33	7.2	3081	4	US-09-762-724-9	Sequence 9, Appli
	6	31.2	6.8	395	3	US-09-328-111-518	Sequence 518, App
	7	31.2	6.8	1939	4	US-08-961-527-310	Sequence 310, App
c	8	30.8	6.8	930	4	US-09-489-039A-3411	Sequence 3411, Ap
	9	30.8	6.8	7766	4	US-09-125-619-3	Sequence 3, Appli
	10	30.6	6.7	1047	3	US-09-315-794-41	Sequence 41, Appl
	11	30.6	6.7	1047	3	US-09-389-341-41	Sequence 41, Appl
c	12	30.6	6.7	1827	4	US-08-956-171E-450	Sequence 450, App
c	13	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	14	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	15	30.4	6.7	7125	1	US-07-745-206A-1	Sequence 1, Appli
	16	30.4	6.7	7125	2	US-08-311-363-1	Sequence 1, Appli
	17	30.4	6.7	7635	1	US-08-455-543A-1	Sequence 1, Appli
	18	30.4	6.7	7635	1	US-08-455-543A-23	Sequence 23, Appl
	19	30.4	6.7	7635	2	US-08-193-078B-1	Sequence 1, Appli
	20	30.4	6.7	7635	2	US-08-193-078B-29	Sequence 29, Appl
	21	30.4	6.7	7635	2	US-08-223-305C-1	Sequence 1, Appli
	22	30.4	6.7	7635	2	US-08-223-305C-23	Sequence 23, Appl
	23	30.4	6.7	7635	2	US-08-149-097D-1	Sequence 1, Appli
	24	30.4	6.7	7635	3	US-08-949-386-1	Sequence 1, Appli
	25	30.4	6.7	7635	3	US-08-450-562-1	Sequence 1, Appli
	26	30.4	6.7	7635	4	US-08-984-709A-1	Sequence 1, Appli
	27	30.4	6.7	7635	4	US-08-450-272-1	Sequence 1, Appli
	28	30.4	6.7	7635	4	US-08-450-273-1	Sequence 1, Appli
	29	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	30	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	31	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
c	32	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	33	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	34	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	35	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
c	36	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
	37	29.4	6.4	3906	2	US-08-469-537A-102	Sequence 102, App
	38	29.4	6.4	4165	1	US-08-442-248-1	Sequence 1, Appli
	39	29.4	6.4	4165	1	US-08-440-815-1	Sequence 1, Appli
	40	29.4	6.4	4165	3	US-08-486-449-1	Sequence 1, Appli
	41	29.4	6.4	4165	4	US-08-578-684-1	Sequence 1, Appli
	42	29.4	6.4	4322	1	US-08-673-789-1	Sequence 1, Appli
	43	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
c	44	29.2	6.4	1491	6	5486473-3	Patent No. 5486473
	45	29.2	6.4	2210	4	US-09-016-434-1177	Sequence 1177, Ap

ALIGNMENTS

```
RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-788A-58

  Query Match             88.8%;  Score 404.8;  DB 1;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.7e-131;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||||||||||||||||||||||||||||||||  |||  ||  ||   |  |||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||  |   ||  |  |  ||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             61.5%;  Score 280.6;  DB 1;  Length 361;
  Best Local Similarity   89.9%;  Pred. No. 3.8e-88;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |||||  |||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 363
              ||||||||||||||||||||||||||||||  |||  ||  ||   |  ||  ||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 GAAATGGCACATGCAAATCAGACCGCATCTGATTC 398
              |  || |  |  ||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

```
RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
;  GENERAL INFORMATION:
;    APPLICANT:  Kim, Jihyun Francis
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;    TITLE OF INVENTION:  ERWINIA AMYLOVORA AND ITS USE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,927
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,108
;      FILING DATE:  06-AUG-1977
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1581
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
```

```
;       TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1344 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.12;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.12;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-762-724-9
; Sequence 9, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
;  APPLICANT: Kovacs, et al.
;  TITLE OF INVENTION: Identification of a region of the major surface
;  TITLE OF INVENTION:  glycoprotein (MSG) gene of human Pneumocystis carinii
;  FILE REFERENCE: 4239-58054
;  CURRENT APPLICATION NUMBER: US/09/762,724
;  CURRENT FILING DATE:  2001-02-09
;  PRIOR APPLICATION NUMBER: PCT/US99/18750
;  PRIOR FILING DATE: 1999-08-17
;  PRIOR APPLICATION NUMBER: US 60/096,805
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3081
;   TYPE: DNA
;   ORGANISM: Pneumocystis carinii sp. f. hominis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3030)
US-09-762-724-9

  Query Match             7.2%;  Score 33;  DB 4;  Length 3081;
  Best Local Similarity   53.5%;  Pred. No. 0.44;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACG 322
              || |||| |  | | |  |||||||||| |||||  |    |  |   |  | ||| |
Db       1395 GTTCTATGGTTCAGGCAAAGAATGGTTTAAGAAACTACTAGAAAAAATAATGGAAGAATG 1454

Qy        323 CTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAATC 382
              |   ||||  | || |  |  |   ||  | ||  | ||| | ||   ||   ||| |
Db       1455 TTCGGAACTTAAAACAACAAGCGATGAGTTGTTTTTGCTATGTATTGATCCACTTAAAGC 1514

Qy        383 AGACCGCAT 391
              || | | ||
Db       1515 AGTCAGAAT 1523


RESULT 6
US-09-328-111-518
; Sequence 518, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: US 60/088,801
;  EARLIER FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 518
;   LENGTH: 395
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-328-111-518

  Query Match             6.8%;  Score 31.2;  DB 3;  Length 395;
  Best Local Similarity   46.7%;  Pred. No. 0.6;
  Matches   99;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
```

```
Db         34 GCAAATGAAGAATGGGAAACAGCTTCTGAAAGCAGTGATTTCAATGAGAGGCGAGAGAGG 93

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTT 354
Db         94 GATGAAAAAAAAAATGCTGACTTGAATGCACAAACAGTTGTAAAGGTTGGAGAGAATGTT 153

Qy        355 GTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
Db        154 CTACCTCCAAAGAGGGAAATTGCAAAGAGAAGTTTTTCTAGTCAGAGACCAGTAGATCGT 213

Qy        415 CAGGTTGGTTTTGGCAACAACGCCACGGCTAA 446
Db        214 CAGAATCGACGTGGCAACAATGGTCCACCCAA 245
```

RESULT 7
US-08-961-527-310
; Sequence 310, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 310:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1939 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-310

Query Match 6.8%; Score 31.2; DB 4; Length 1939;
Best Local Similarity 51.4%; Pred. No. 1.4;
Matches 72; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
Db        618 GCCAATGTAGTCACGGTTACGGAAAAAGACTTTACTGAATTTCCTAAGAGTGACACCAGT 677

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTT 354
Db        678 AAGGTTCCTATCCTAGATAGAAGTACTGCTGAAAAAATTGGAGACCGCTACTTGGGTTCC 737

Qy        355 GTTACCCATGAAATGGCACA 374
Db        738 CTAACCGATAAGGTGTCGCA 757
```

RESULT 8
US-09-489-039A-3411/c
; Sequence 3411, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3411
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3411

Query Match 6.8%; Score 30.8; DB 4; Length 930;
Best Local Similarity 46.3%; Pred. No. 1.3;
Matches 101; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

```
Qy        213 CCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGA 272
Db        262 CCCCGGCGGTCAGGTAGATCGCCATGGCGACGCCCAGACCCCAGATGATGCTGATTTCCC 203

Qy        273 ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTA 332
Db        202 ATTGTCCGAAGCTGGCTCCCGCGACCTTGAGCGCAGCCACGCACCCAACGCCGAAAAAGA 143

Qy        333 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATC 392
Db        142 TCAACAACCCGGTACCGAGGAACTCTGCGATGCACTGGCCTTTTAAGGTTGATGTTTGGC 83

Qy        393 TGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCA 430
Db         82 TCATAATCGAATCCTGAAGCGTGGTGATGATTATTGTA 45
```

RESULT 9
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234
; CURRENT APPLICATION NUMBER: US/09/125,619
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match 6.8%; Score 30.8; DB 4; Length 7766;
Best Local Similarity 48.8%; Pred. No. 4.3;
Matches 83; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              |  ||  ||||||  ||||     |    ||||     ||  |||  |  |  |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  |  |  ||  ||  ||    | |   |   | |||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |   ||  |||  |    |    |  |||  |           | |||  |||||
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAA 4601


RESULT 10
US-09-315-794-41
; Sequence 41, Application US/09315794
; Patent No. 6197517
; GENERAL INFORMATION:
;  APPLICANT: Roberts, Christopher J.
;  TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
;  TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
;  TITLE OF INVENTION: DRUGS
;  FILE REFERENCE: 9301-053
;  CURRENT APPLICATION NUMBER: US/09/315,794
;  CURRENT FILING DATE: 1999-05-21
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
;   LENGTH: 1047
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-315-794-41

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 1047;
  Best Local Similarity   55.0%;  Pred. No. 1.7;
  Matches   60;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        271 GAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAAC 330
              || |||| | || || |         || ||  | | |  |||   |  || ||  || |
Db        601 GACCTGATTGAGGATAGCGATGATGATGATTTCTCTAGTGACTCTTCCAAAGCCCAATAT 660

Qy        331 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAA 379
              | ||||||  ||| |||| | | || | ||  ||||   | | |   ||
Db        661 TTTGATCAAGTGGATACCAGCGATGATTCCATTGAAGAAGAAGAATTAA 709


RESULT 11
US-09-389-341-41
; Sequence 41, Application US/09389341
; Patent No. 6200803
; GENERAL INFORMATION:
;  APPLICANT: Roberts, Christopher J.
;  TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
;  TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATIVE
;  TITLE OF INVENTION: DRUGS
;  FILE REFERENCE: 9301-057
;  CURRENT APPLICATION NUMBER: US/09/389,341
;  CURRENT FILING DATE: 1999-09-02
;  EARLIER APPLICATION NUMBER: 09/315,794
;  EARLIER FILING DATE: 1999-05-21
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
;   LENGTH: 1047
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-389-341-41

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 1047;
  Best Local Similarity   55.0%;  Pred. No. 1.7;
  Matches   60;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        271 GAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAAC 330
              || |||| | || || |         || ||  | | |  |||   |  || ||  || |
Db        601 GACCTGATTGAGGATAGCGATGATGATGATTTCTCTAGTGACTCTTCCAAAGCCCAATAT 660

Qy        331 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAA 379
              | ||||||  ||| |||| | | || | ||  ||||   | | |   ||
Db        661 TTTGATCAAGTGGATACCAGCGATGATTCCATTGAAGAAGAAGAATTAA 709


RESULT 12
US-08-956-171E-450/c
; Sequence 450, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/956,171E
;             FILING DATE: 20-Oct-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 450:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1827 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 450:
US-08-956-171E-450

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 1827;
  Best Local Similarity   47.6%;  Pred. No. 2.3;
  Matches   90;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        150 CGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCAT 209
              ||   | | |||   ||  |  ||   |   | | ||      |||   | ||    |||
Db        710 CGCAGCAGATAATTTTGGTCGAGCAGCGGCTATCAATCAAGTAAAAAAAGTAATCTACAT 651

Qy        210 TACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTAT 269
                |   |||  |||||| ||| |   |   || | |||  |||||  ||  ||   || |
Db        650 ACCTGGGAGTCGTTATGATAATGAAACAATTGAACGCCTAGGTGCATATGGTACACCTGT 591

Qy        270 TGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAA 329
              |||   ||  |   | ||| | |  |  |      |  |   | |   | | |  |||
```

```
Db        590 AGAAACAACAAATTTAGTTTTTAAACGTTCTTTAGTTAATGTAGAATTACAAGTTTCAAA 531

Qy        330 CTATGATCA 338
              |||||| |
Db        530 GTATGATGA 522


RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4403765;
  Best Local Similarity   51.9%;  Pred. No. 94;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || ||||||| || ||  |  ||   | |||||| |||| |  |   || ||
Db    1357218 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357159

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
              ||     | | |     ||  | ||| | ||    |   ||   ||  || | |  |||
Db    1357158 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357099

Qy        165 TGCGCTTGCTCTG 177
              ||   |  |||
Db    1357098 CCCGGACGTGCTG 1357086


RESULT 14
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4411529;
  Best Local Similarity   51.9%;  Pred. No. 94;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || ||||||| || ||  |  ||   | |||||| |||| |  |   || ||
Db    1357750 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357691

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
              ||     | | |     ||  | ||| | ||    |   ||   ||  || | |  |||
Db    1357690 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357631

Qy        165 TGCGCTTGCTCTG 177
              ||   |  |||
Db    1357630 CCCGGACGTGCTG 1357618


RESULT 15
US-07-745-206A-1
; Sequence 1, Application US/07745206A
; Patent No. 5429921
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Feldman, Daniel
;    TITLE OF INVENTION:  Human Calcium Channel Compositions and
;    TITLE OF INVENTION:  Methods
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fitch, Even, Tabin & Flannery
;      STREET:  135 S. LaSalle
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/745,206A
;      FILING DATE:  19910815
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feder, Scott B
;      REFERENCE/DOCKET NUMBER:  51504
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-372-7842
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7125 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..6483
US-07-745-206A-1

  Query Match             6.7%;  Score 30.4;  DB 1;  Length 7125;
  Best Local Similarity   57.3%;  Pred. No. 5.7;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              | |||   ||| |  |   |  ||| |||| | ||||   ||||  | |    ||| |
Db       6839 CTGAAGGCATGGCGGCGGGGTGCAGGGGAAAGTTAAAGGTGATGACGATCATCACACCTG 6898

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGAC 386
              ||| ||||||   |  || |  |  ||| ||||| |
```

Db 6899 TGTCGTTACCTCAGCCATCGGTCTAGCATATCAGTC 6934

Search completed: March 15, 2004, 10:14:45
Job time : 68.8 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 07:24:24 ; Search time 340.6 Seconds
(without alignments)
4928.974 Million cell updates/sec

Title: US-09-543-407-23
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	37	8.1	2374	14	US-10-237-271-6	Sequence 6, Appli
c 2	37	8.1	2374	15	US-10-353-690-125	Sequence 125, App
c 3	37	8.1	2383	13	US-10-044-090-521	Sequence 521, App
c 4	34.2	7.5	616	12	US-10-424-599-88945	Sequence 88945, A
c 5	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c 6	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c 7	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c 8	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c 9	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
c 10	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
11	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
12	33	7.2	3081	12	US-10-654-416-9	Sequence 9, Appli
13	32.6	7.1	14036	10	US-09-764-872-818	Sequence 818, App
14	32.2	7.1	258	9	US-09-783-590-8384	Sequence 8384, Ap
15	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
16	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c 17	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
18	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
19	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
20	31.6	6.9	549	13	US-10-040-739-702	Sequence 702, App
c 21	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
22	31.2	6.8	395	9	US-09-879-536-518	Sequence 518, App
23	31.2	6.8	1374	9	US-09-864-761-19619	Sequence 19619, A
24	31.2	6.8	1939	12	US-10-158-844-310	Sequence 310, App
25	31.2	6.8	1965	9	US-09-864-761-2837	Sequence 2837, Ap
26	31.2	6.8	2886	9	US-09-801-368-131	Sequence 131, App
27	31.2	6.8	10174	14	US-10-171-311-82	Sequence 82, Appl
c 28	31	6.8	2226	12	US-10-425-114-28851	Sequence 28851, A
c 29	30.8	6.8	441	9	US-09-864-761-3403	Sequence 3403, Ap
c 30	30.8	6.8	1977	12	US-10-282-122A-28900	Sequence 28900, A
31	30.8	6.8	7766	12	US-10-222-566-3	Sequence 3, Appli
32	30.8	6.8	7766	14	US-10-222-162-3	Sequence 3, Appli
33	30.8	6.8	7766	14	US-10-143-024-3	Sequence 3, Appli
34	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
c 35	30.6	6.7	1827	8	US-08-781-986A-450	Sequence 450, App
c 36	30.6	6.7	1827	12	US-10-329-624-450	Sequence 450, App
37	30.6	6.7	2980	12	US-10-424-599-1292	Sequence 1292, Ap
c 38	30.6	6.7	75216	14	US-10-080-170-646	Sequence 646, App
39	30.4	6.7	545	14	US-10-060-036-1674	Sequence 1674, Ap
c 40	30.4	6.7	992	12	US-10-425-114-31747	Sequence 31747, A
c 41	30.4	6.7	6907	14	US-10-272-715-1	Sequence 1, Appli
42	30.4	6.7	7362	13	US-10-029-413A-15	Sequence 15, Appl
43	30.4	6.7	7635	15	US-10-375-253-1	Sequence 1, Appli
44	30.2	6.6	410	9	US-09-974-300-6264	Sequence 6264, Ap
45	30.2	6.6	444	9	US-09-960-352-14115	Sequence 14115, A

ALIGNMENTS

RESULT 1
US-10-237-271-6/c
; Sequence 6, Application US/10237271
; Publication No. US20030096328A1
; GENERAL INFORMATION:
; APPLICANT: THE BURNHAM INSTITUTE
; APPLICANT: SMITH, Jeffrey W.
; APPLICANT: KRIDEL, Steven J.
; APPLICANT: AXELROD, Fumiko T.
; TITLE OF INVENTION: SERINE/THREONINE HYDROLASE PROTEINS AND SCREENING ASSAYS
; FILE REFERENCE: BURN1100-1
; CURRENT APPLICATION NUMBER: US/10/237,271
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 60/317,842
; PRIOR FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 2374
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-237-271-6

Query Match 8.1%; Score 37; DB 14; Length 2374;
Best Local Similarity 52.2%; Pred. No. 0.11;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy      291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
            |||| ||  |||||  |||||     | ||||  |  ||   |||    | |  | ||
Db      725 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 666

Qy      351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
            |||| |  |||  |   | | | |    || | || | |  ||||      ||   | |
Db      665 TGTTTTCTCCCCAGTCTTCGTAAAACACAAACTGATCCCCCTTGATGGGTTGGTCTGGCT 606

Qy      411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              ||||  ||| | |  || || |||     || ||
Db      605 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 569
```

```
RESULT 2
US-10-353-690-125/c
; Sequence 125, Application US/10353690
; Publication No. US20030215840A1
; GENERAL INFORMATION:
;  APPLICANT: Logan, Thomas Joseph
;  APPLICANT:  Chun, Miyoung
;  APPLICANT:  Galvin, Katherine M.
;  APPLICANT:  Healy, Aileen
;  APPLICANT:  Acton, Susan L.
;  APPLICANT:  Donoghue, Mary
;  APPLICANT:  Stagliano, Nancy
;  APPLICANT:  Perodin, Jacquelin
;  APPLICANT:  Rodrigue-Way, Amelie
;  TITLE OF INVENTION: Methods and compositions for treating
;  TITLE OF INVENTION:  cardiovascular disease using 1682, 6169, 6193, 7771, 14395,
;  TITLE OF INVENTION:  29002, 33216, 43726, 69292, 26156, 32427, 2402, 7747, 1720,
;  TITLE OF INVENTION:  9151, 60491, 1371, 7077, 33207, 1419, 18036, 16105, 38650,
;  TITLE OF INVENTION:  14245, 58848, 1870, 25856, 32394, 3484, 345, 9252, 9135,
;  TITLE OF INVENTION:  10532, 18610, 8165, 2448, 2445, 64624, 84237, 8912, 2868,
;  TITLE OF INVENTION:  283, 2554, 9464, 17799, 26686, 43848, 32135, 12208, 2914,
;  TITLE OF INVENTION:  51130, 19489, 21833, 2917, 59590, 15992, 2094, 2252, 3474,
;  TITLE OF INVENTION:  9792, 15400, 1452 or 6585 molecules
;  FILE REFERENCE: MPI02-018P1RNOMNIM
;  CURRENT APPLICATION NUMBER: US/10/353,690
;  CURRENT FILING DATE:  2003-01-29
;  PRIOR APPLICATION NUMBER: 60/353,224
;  PRIOR FILING DATE: 2002-02-01
;  PRIOR APPLICATION NUMBER: 60/364,529
;  PRIOR FILING DATE: 2002-03-15
;  PRIOR APPLICATION NUMBER: 60/373,861
;  PRIOR FILING DATE: 2002-04-19
;  PRIOR APPLICATION NUMBER: 60/376,287
;  PRIOR FILING DATE: 2002-04-29
;  PRIOR APPLICATION NUMBER: 60/388,080
;  PRIOR FILING DATE: 2002-06-12
;  PRIOR APPLICATION NUMBER: 60/390,971
;  PRIOR FILING DATE: 2002-06-24
;  PRIOR APPLICATION NUMBER: 60/394,130
;  PRIOR FILING DATE: 2002-07-03
;  PRIOR APPLICATION NUMBER: 60/394,797
;  PRIOR FILING DATE: 2002-07-10
;  PRIOR APPLICATION NUMBER: 60/404,904
;  PRIOR FILING DATE: 2002-08-21
;  PRIOR APPLICATION NUMBER: 60/405,450
;  PRIOR FILING DATE: 2002-08-23
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 126
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 125
;   LENGTH: 2374
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
US-10-353-690-125

  Query Match             8.1%;  Score 37;  DB 15;  Length 2374;
  Best Local Similarity   52.2%;  Pred. No. 0.11;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        725 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 666

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              |||| |  |||  |   | | | |    || | || | |   ||||       ||   | |
Db        665 TGTTTTCTCCCCAGTCTTCGTAAAACACAAACTGATCCCCCTTGATGGGTTGGTCTGGCT 606

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  |||  | |  || || |||     || ||
Db        605 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 569
```

```
RESULT 3
US-10-044-090-521/c
; Sequence 521, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 521
;   LENGTH: 2383
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 3463621CB1
US-10-044-090-521

  Query Match             8.1%;  Score 37;  DB 13;  Length 2383;
  Best Local Similarity   52.2%;  Pred. No. 0.11;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        719 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 660

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              |||| |  |||  |   | | | |    || | || | |   ||||       ||   | |
Db        659 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 600

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  |||  | |  || || |||     || ||
Db        599 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 563
```

```
RESULT 4
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
;   LENGTH: 616
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(616)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 616;
  Best Local Similarity   50.9%;  Pred. No. 0.5;
  Matches   81;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
              | | |  |||  ||  || ||| || || |||  | |    | |  | | | |   | |
Db        556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497
```

```
Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              | | ||    || |  | |    || ||  |  |   | | |||   ||   |  | |
Db        496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437

Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              || ||||||  |  |   ||| |||  ||||| ||| ||
Db        436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTA 398
```

RESULT 5
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

```
  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.91;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 6
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

```
  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.91;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 7
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

```
  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.91;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 8
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 6
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-010-390-6

  Query Match             7.5%;  Score 34;  DB 14;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.91;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 9
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION:  RESISTANCE
;  FILE REFERENCE: 21829/203 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/10/441,736
;  CURRENT FILING DATE:  2003-05-20
;  PRIOR APPLICATION NUMBER: 60/107,243
;  PRIOR FILING DATE: 1998-11-05
;  PRIOR APPLICATION NUMBER: 09/431,614
;  PRIOR FILING DATE: 1999-11-02
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-441-736-5

  Query Match             7.5%;  Score 34;  DB 15;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.91;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 10
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
;   LENGTH: 1662
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1662)
US-10-156-761-3483

  Query Match             7.3%;  Score 33.2;  DB 14;  Length 1662;
  Best Local Similarity   52.1%;  Pred. No. 1.9;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
              | || | ||  | ||||| | ||| |  | | || || || | ||| | ||||||   |
Db        590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531

Qy        110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
                 | ||| ||  | | |   ||   ||  | |     |  || |||   | | |||
Db        530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

Qy        170 TTGCTCTGCAAAGCGATGCCCG 191
              |||  || |  |||  | | |
Db        470 TTGAACTTCTTGGCGGCGACGG 449


RESULT 11
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             7.3%;  Score 33.2;  DB 14;  Length 9025608;
  Best Local Similarity   52.1%;  Pred. No. 1.9e+02;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;
```

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               |||| | ||  | ||||| | || |  | | || || || | ||| | ||||||   |
Db     4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

Qy         110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
                 | ||| ||  | | |   ||   ||  | |      |  || |||   | | |||
Db     4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

Qy         170 TTGCTCTGCAAAGCGATGCCCG 191
               |||  || |  |||  | | |
Db     4331901 TTGAACTTCTTGGCGGCGACGG 4331922


RESULT 12
US-10-654-416-9
; Sequence 9, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
;  APPLICANT: Kovacs, et al.
;  TITLE OF INVENTION: Identification of a region of the major surface
;  TITLE OF INVENTION:  glycoprotein (MSG) gene of human Pneumocystis carinii
;  FILE REFERENCE: 4239-58054
;  CURRENT APPLICATION NUMBER: US/10/654,416
;  CURRENT FILING DATE:  2003-09-02
;  PRIOR APPLICATION NUMBER: US/09/762,724
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: PCT/US99/18750
;  PRIOR FILING DATE: 1999-08-17
;  PRIOR APPLICATION NUMBER: US 60/096,805
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3081
;   TYPE: DNA
;   ORGANISM: Pneumocystis carinii sp. f. hominis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3030)
US-10-654-416-9

  Query Match             7.2%;  Score 33;  DB 12;  Length 3081;
  Best Local Similarity   53.5%;  Pred. No. 3.2;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy         263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACG 322
               || |||| |  | | |  |||||||||| |||||  |    |  |    |  |||  |
Db        1395 GTTCTATGGTTCAGGCAAAGAATGGTTTAAGAAACTACTAGAAAAAATAATGGAAGAATG 1454

Qy         323 CTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAATC 382
                 |   ||||   | || |  |  |   ||  | ||  | ||| | ||   ||   ||| |
Db        1455 TTCGGAACTTAAAACAACAAGCGATGAGTTGTTTTTGCTATGTATTGATCCACTTAAAGC 1514

Qy         383 AGACCGCAT 391
               || | | ||
Db        1515 AGTCAGAAT 1523


RESULT 13
US-09-764-872-818
; Sequence 818, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA125
;  CURRENT APPLICATION NUMBER: US/09/764,872
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 957
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 818
;   LENGTH: 14036
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-872-818

  Query Match             7.1%;  Score 32.6;  DB 10;  Length 14036;
  Best Local Similarity   60.9%;  Pred. No. 11;
  Matches   53;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         336 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGA 395
               ||||||| || | | |    |  || | |||||  |  |  ||||||| | |  |
Db        2638 TCAGCTGTTTGCTCATCTGAACAGGCAGGGAATGGTCCTAGATAATCAGATCTCTCCGAC 2697

Qy         396 TTCCAGCGTAATGGTGCGTCAGGTTGG 422
               | ||||| |  ||  ||| || |||||
Db        2698 TGCCAGCATTCTGTGGCGCCATGTTGG 2724


RESULT 14
US-09-783-590-8384
; Sequence 8384, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,731
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8384
;   LENGTH: 258
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (10)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (49)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (130)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (147)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (201)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (211)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (247)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-8384

  Query Match             7.1%;  Score 32.2;  DB 9;  Length 258;
  Best Local Similarity   52.3%;  Pred. No. 1.5;
  Matches   67;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         252 TGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGA 311
               ||   | ||   |||   | |||| || |  ||||||  |||     | |  |||
Db          76 TGTTCAAAAGCAGACTCGGGGACTGCCTGAATGTGGTTTGGGAACACGAGACCTNATCAT 135
```

```
Qy        312 CCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGC 371
              |  || ||  | |||| | || ||  | | | | |    |  ||      |||||||||
Db        136 ATTGATGAGCGNAACAAACAAGGAACATTTCCTCCCTCCCCTCCTTTGAATTGAAATGGC 195

Qy        372 ACATGCAA 379
              ||||  ||
Db        196 ACATTNAA 203


RESULT 15
US-10-333-631-1
; Sequence 1, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer AG
;  TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
;  TITLE OF INVENTION:  ENZYME
;  FILE REFERENCE: LIO121 Foreign Countries
;  CURRENT APPLICATION NUMBER: US/10/333,631
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US 60/221,917
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/280,112
;  PRIOR FILING DATE: 2001-04-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 568
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90)..(90)
;   OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

  Query Match             7.1%;  Score 32.2;  DB 14;  Length 568;
  Best Local Similarity   52.6%;  Pred. No. 2.4;
  Matches   70;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCC 214
              | ||| |    |  ||||| |||||   ||   || ||    |||||| |  |  |  ||
Db        103 CTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCTGAAATGGTCCGTCACC 162

Qy        215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              ||  |    || || | || | || | ||| ||||||||    | |  ||   ||
Db        163 CAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTACCTGGATCTGCGGATA 222

Qy        275 TGACTCAGAATGG 287
              | ||   |||
Db        223 GCCCACATGCTGG 235


Search completed: March 15, 2004, 14:16:18
Job time : 352.6 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:52:04 ; Search time 2500.2 Seconds
(without alignments)
6417.435 Million cell updates/sec

Title: US-09-543-407-23
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-23	Sequence 23, Appl
2	404.8	88.8	456	6	US-08-233-642A-56	Sequence 56, Appl
3	404.8	88.8	456	23	US-09-543-407-1	Sequence 1, Appli
4	382.4	83.9	456	23	US-09-543-407-13	Sequence 13, Appl
5	377.6	82.8	456	23	US-09-543-407-29	Sequence 29, Appl
6	361.6	79.3	456	23	US-09-543-407-19	Sequence 19, Appl
7	355.2	77.9	456	23	US-09-543-407-17	Sequence 17, Appl
8	353.6	77.5	456	23	US-09-543-407-11	Sequence 11, Appl
9	353.6	77.5	456	23	US-09-543-407-25	Sequence 25, Appl
10	352	77.2	456	23	US-09-543-407-15	Sequence 15, Appl
11	350.4	76.8	456	23	US-09-543-407-21	Sequence 21, Appl
12	350.4	76.8	456	23	US-09-543-407-27	Sequence 27, Appl
13	280.6	61.5	361	6	US-08-233-642A-54	Sequence 54, Appl
14	238.4	52.3	456	23	US-09-543-407-3	Sequence 3, Appli
15	238.4	52.3	675	13	US-08-978-878-3	Sequence 3, Appli
16	238.4	52.3	675	31	US-09-741-873B-3	Sequence 3, Appli
17	236	51.8	453	92	US-60-352-946-1	Sequence 1, Appli
18	236	51.8	453	101	US-60-444-371-1	Sequence 1, Appli
19	233.6	51.2	477	17	US-09-252-691-172	Sequence 172, App
20	233.6	51.2	477	17	US-09-252-691C-172	Sequence 172, App
21	233.6	51.2	477	51	US-10-417-886-172	Sequence 172, App
22	194.4	42.6	396	13	US-08-978-878-1	Sequence 1, Appli
23	169.6	37.2	360	31	US-09-741-873B-1	Sequence 1, Appli
c 24	78	17.1	78	23	US-09-543-407-52	Sequence 52, Appl
25	78	17.1	78	23	US-09-543-407-53	Sequence 53, Appl
26	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
27	44.2	9.7	522	17	US-09-252-691-170	Sequence 170, App
28	44.2	9.7	522	17	US-09-252-691C-170	Sequence 170, App
29	44.2	9.7	522	51	US-10-417-886-170	Sequence 170, App
30	41	9.0	100	45	US-10-146-492B-77	Sequence 77, Appl
31	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
c 32	40.6	8.9	78	23	US-09-543-407-48	Sequence 48, Appl
33	40.6	8.9	78	23	US-09-543-407-49	Sequence 49, Appl
c 34	40	8.8	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
35	37.4	8.2	1304	51	US-10-437-963-19773	Sequence 19773, A
c 36	37	8.1	292	18	US-09-332-782-7392	Sequence 7392, Ap
c 37	37	8.1	292	21	US-09-515-694-7392	Sequence 7392, Ap
c 38	37	8.1	998	19	US-09-371-168-6052	Sequence 6052, Ap
c 39	37	8.1	998	27	US-09-644-870-7155	Sequence 7155, Ap
c 40	37	8.1	998	27	US-09-652-127-8932	Sequence 8932, Ap
c 41	37	8.1	998	27	US-09-652-355-8355	Sequence 8355, Ap
c 42	37	8.1	998	27	US-09-652-816-6743	Sequence 6743, Ap
c 43	37	8.1	998	33	US-09-801-833-6052	Sequence 6052, Ap
c 44	37	8.1	2188	77	US-60-213-172-318	Sequence 318, App
c 45	37	8.1	2188	77	US-60-213-173-318	Sequence 318, App

ALIGNMENTS

```
RESULT 1
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 5.2e-135;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.4e-118;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||     ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.4e-118;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||     ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             83.9%;  Score 382.4;  DB 23;  Length 456;
  Best Local Similarity   89.9%;  Pred. No. 2.2e-111;
  Matches  410;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 82.8%; Score 377.6; DB 23; Length 456;
Best Local Similarity 89.3%; Pred. No. 7.6e-110;
Matches 407; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
Db        301 GCCACCATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 1e-104;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 1.2e-102;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||      ||   ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 3.9e-102;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||  |||      ||   ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               || |    || |  |  ||| |           | | || | ||  |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 3.9e-102;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||  |    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||    |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||  |||     ||   ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 1.3e-101;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||     | | | |    | |      |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              |||||||||||||||||||||||||||||||  |||     ||   ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 4.1e-101;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |  |      | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||       ||||||||||||||||||||||  |||      ||    ||      |
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match            76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 4.1e-101;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||  |      |  |  ||   |     |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              ||||||||||||||||||||||||||||||||  |||      ||    ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-642A-54

  Query Match            61.5%;  Score 280.6;  DB 6;  Length 361;
  Best Local Similarity   89.9%;  Pred. No. 9.5e-79;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 363
              |||||||||||||||||||||||||||  |||     ||   ||     |     ||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 GAAATGGCACATGCAAATCAGACCGCATCTGATTC 398
              |   ||| |  ||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-3

  Query Match             52.3%;  Score 238.4;  DB 23;  Length 456;
  Best Local Similarity   70.2%;  Pred. No. 3.6e-65;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||  |||| | |||||||||||| | || || || |||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || ||  | || ||||||||||  ||||| | |||  |  |||||  |  | | |
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |    ||   |  | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

```
RESULT 15
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
;  APPLICANT: NORMARK, Staffan
;  APPLICANT: OLSEN, Arne
;  TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
;  FILE REFERENCE: 012889-081
;  CURRENT APPLICATION NUMBER: US/08/978,878
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: SE 8801723-1
;  EARLIER FILING DATE: 1988-05-06
;  EARLIER APPLICATION NUMBER: US 07/347,189
;  EARLIER FILING DATE: 1989-05-04
;  EARLIER APPLICATION NUMBER: US 07/789,437
;  EARLIER FILING DATE: 1991-11-06
;  EARLIER APPLICATION NUMBER: US 07/970,846
;  EARLIER FILING DATE: 1992-11-03
;  EARLIER APPLICATION NUMBER: US 08/187,865
;  EARLIER FILING DATE: 1994-01-28
;  EARLIER APPLICATION NUMBER: US 08/318,519
;  EARLIER FILING DATE: 1994-10-05
;  EARLIER APPLICATION NUMBER: US 08/495,959
;  EARLIER FILING DATE: 1995-06-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-08-978-878-3

  Query Match             52.3%;  Score 238.4;  DB 13;  Length 675;
  Best Local Similarity   70.2%;  Pred. No. 4.2e-65;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||  |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || ||  | || ||||||||||  ||||| | |||  |  |||||  |  | | |
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |    ||   |  | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

Search completed: March 15, 2004, 13:42:48
Job time : 2501.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 04:16:29 ; Search time 53.4 Seconds
(without alignments)
2945.701 Million cell updates/sec

Title: US-09-543-407-23
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 628836 seqs, 172478556 residues

Total number of hits satisfying chosen parameters: 1257672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	238.4	52.3	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	194.4	42.6	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
	4	31.4	6.9	1588	6	US-10-767-701-14522	Sequence 14522, A
	5	30.8	6.8	3990	5	US-09-830-230A-629	Sequence 629, App
c	6	30	6.6	24900	6	US-10-767-471-10736	Sequence 10736, A
c	7	29.8	6.5	261922	6	US-10-767-471-10653	Sequence 10653, A
	8	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
	9	29.4	6.4	1355	6	US-10-045-674A-593	Sequence 593, App
	10	29.2	6.4	600	7	US-60-545-213-2020	Sequence 2020, Ap
	11	29.2	6.4	600	7	US-60-545-213-6292	Sequence 6292, Ap
	12	28.6	6.3	921	6	US-10-767-795-3876	Sequence 3876, Ap
	13	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	14	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	15	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	16	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
	17	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
	18	28	6.1	1870	6	US-10-645-546-3	Sequence 3, Appli
	19	28	6.1	3707	6	US-10-645-546-1	Sequence 1, Appli
	20	28	6.1	8192	6	US-10-645-546-5	Sequence 5, Appli
	21	28	6.1	8240	6	US-10-645-546-7	Sequence 7, Appli
	22	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
c	23	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	24	27.6	6.1	614	6	US-10-100-683-3983	Sequence 3983, Ap
	25	27.6	6.1	774	6	US-10-100-683-11482	Sequence 11482, A
	26	27.6	6.1	791	6	US-10-100-683-900	Sequence 900, App
	27	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	28	27.6	6.1	1248	6	US-10-767-795-1845	Sequence 1845, Ap
	29	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	30	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	31	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	32	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	33	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
	34	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
	35	27.4	6.0	261922	6	US-10-767-471-10653	Sequence 10653, A
	36	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
c	37	27.2	6.0	611	6	US-10-767-701-4907	Sequence 4907, Ap
	38	27.2	6.0	772	6	US-10-767-701-9549	Sequence 9549, Ap
	39	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
c	40	27.2	6.0	2880	6	US-10-100-683-5057	Sequence 5057, Ap
c	41	27.2	6.0	2881	6	US-10-100-683-2682	Sequence 2682, Ap
c	42	27.2	6.0	7380	6	US-10-417-375A-23	Sequence 23, Appl
c	43	27.2	6.0	7400	6	US-10-417-375A-19	Sequence 19, Appl
c	44	27.2	6.0	7460	6	US-10-417-375A-25	Sequence 25, Appl
c	45	27.2	6.0	113486	6	US-10-417-375A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 52.3%; Score 238.4; DB 5; Length 675;
Best Local Similarity 70.2%; Pred. No. 7.2e-72;
Matches 320; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 42.6%; Score 194.4; DB 5; Length 396;
Best Local Similarity 68.2%; Pred. No. 6.4e-57;
Matches 270; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
; APPLICANT: Neubauer, Antonie
; APPLICANT: Ziegler, Christina
; TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
; FILE REFERENCE: 1/1372
; CURRENT APPLICATION NUMBER: US/10/624,149A
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 60/403,282
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: DE 10233064
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: DE 10317008
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
; LENGTH: 150223
; TYPE: DNA
; ORGANISM: Equine herpesvirus 1
; PUBLICATION INFORMATION:
; AUTHORS: Telford,E.A.
; AUTHORS: Watson, M.S.
; AUTHORS: McBride, K.
; AUTHORS: Davison, A.J.
; TITLE: The DNA sequence of equine herpesvirus-1
; JOURNAL: Virology
; VOLUME: 189
; ISSUE: 1
; PAGES: 304-316
; DATE: JUL-1992
; DATABASE ACCESSION NUMBER: NC 001491, NCBI
; DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 2.9;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232

Db      95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy        233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292

Db      95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy        293 GAAATAATGCCACCATCGACCAGTGG 318

Db      95662 GAGAGCTTATTCACAACGAGCTGTTG 95687
```

RESULT 4
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:

```
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
;   LENGTH: 1588
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 1588;
  Best Local Similarity   44.5%;  Pred. No. 0.77;
  Matches  125;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
              |  |  ||||  |       | ||| | |  |       | |||||  |||  |  |  |  |
Db        455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
              |||||  |||  ||  ||   |||   |     |  |   |||                |||| | |
Db        515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy        160 AACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
                | ||   |||  |    |  | | |||  ||  | |||  |   | | ||  |      |
Db        575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              |||  |   | |   |||          |  |||||  ||  | | ||  |  ||  |    ||
Db        635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAA 320
                | | ||  |   |  | ||||   || | ||  | |
Db        695 GCTGGTAGCTTTGGCGGTGGTTCCACTGGCGCCGAGAAGGA 735


RESULT 5
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Lyme Disease Vaccines
;  FILE REFERENCE: PB481US
;  CURRENT APPLICATION NUMBER: US/09/830,230A
;  CURRENT FILING DATE:  2001-09-27
;  PRIOR APPLICATION NUMBER: PCT/US98/12718
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/057,483
;  PRIOR FILING DATE: 1997-09-03
;  PRIOR APPLICATION NUMBER: 60/053,344
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/053,377
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/050,359
;  PRIOR FILING DATE: 1997-06-20
;  NUMBER OF SEQ ID NOS: 756
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
;   LENGTH: 3990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1139)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1143)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1148)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1210)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1244)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1247)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1250)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1251)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

  Query Match             6.8%;  Score 30.8;  DB 5;  Length 3990;
  Best Local Similarity   48.8%;  Pred. No. 1.9;
  Matches   83;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||     ||  ||| |  | | |  |  | ||
Db       3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |   |  | ||||||||||  | |||    |||  ||
Db       3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |   || ||| |   |    |  ||| |          | |||  |||||
Db       3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAA 3602


RESULT 6
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
;   LENGTH: 24900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(24900)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

  Query Match             6.6%;  Score 30;  DB 6;  Length 24900;
  Best Local Similarity   52.4%;  Pred. No. 8.4;
  Matches   66;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||  | ||  ||  | ||  | ||  |       |   | ||  ||  |  | | |||||  |
```

```
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326

Db       9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929

Qy        327 AAACTA 332

Db       9928 AATCAA 9923
```

RESULT 7
US-10-767-471-10653/c
; Sequence 10653, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10653
; LENGTH: 261922
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(261922)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10653

Query Match 6.5%; Score 29.8; DB 6; Length 261922;
Best Local Similarity 54.0%; Pred. No. 29;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy        306 CATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGA 365

Db     247788 CATCAGACAGCAGAAAACCAGGTCTGAAAATCATCACTTAGCTGGTGTTATGGCTCATGT 247729

Qy        366 AATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGG 418

Db     247728 TATGGCTCAGTTATATCACACTGAAGGTAAAGCTATTAAAATGTTACGGTAGG 247676
```

RESULT 8
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

Query Match 6.4%; Score 29.4; DB 6; Length 300;
Best Local Similarity 54.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73

Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124

Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233
```

RESULT 9
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
US-10-045-674A-593

Query Match 6.4%; Score 29.4; DB 6; Length 1355;
Best Local Similarity 25.2%; Pred. No. 3.5;
Matches 39; Conservative 50; Mismatches 66; Indels 0; Gaps 0;

```
Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137

Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197

Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232

Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYG 922
```

RESULT 10
US-60-545-213-2020
; Sequence 2020, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2020

```
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2020

  Query Match             6.4%;  Score 29.2;  DB 7;  Length 600;
  Best Local Similarity   49.4%;  Pred. No. 2.8;
  Matches   76;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316
              ||| ||   |  |||| | || ||| |||    || |  |    |  || | | ||| |
Db         83 ATATGAGAGATGATGAAATAACCCAGGATGAAAACACAGTTTCCACTTCCCTGGGCCATT 142

Qy        317 GGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 376
              || | | | |||| | | ||       | | |  |||  ||| |       |  ||
Db        143 CCAAAGATGAGAACTCTAAGCAAACATGCATCAGAATTGGCACCAAGACCCCAAAAAGTG 202

Qy        377 CAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
                  ||| || || || ||| |||   |||
Db        203 ACTCATGTACCCCAACTAATACCACCGTGGAGGT 236


RESULT 11
US-60-545-213-6292
; Sequence 6292, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6292
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6292

  Query Match             6.4%;  Score 29.2;  DB 7;  Length 600;
  Best Local Similarity   49.4%;  Pred. No. 2.8;
  Matches   76;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316
              ||| ||   |  |||| | || ||| |||    || |  |    |  || | | ||| |
Db         83 ATATGAGAGATGATGAAATAACCCAGGATGAAAACACAGTTTCCACTTCCCTGGGCCATT 142

Qy        317 GGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 376
              || | | | |||| | | ||       | | |  |||  ||| |       |  ||
Db        143 CCAAAGATGAGAACTCTAAGCAAACATGCATCAGAATTGGCACCAAGACCCCAAAAAGTG 202

Qy        377 CAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
                  ||| || || || ||| |||   |||
Db        203 ACTCATGTACCCCAACTAATACCACCGTGGAGGT 236


RESULT 12
US-10-767-795-3876
; Sequence 3876, Application US/10767795
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Zhou, Yihua
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53534)B
;  CURRENT APPLICATION NUMBER: US/10/767,795
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 3876
;   LENGTH: 921
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C2531_1
US-10-767-795-3876

  Query Match             6.3%;  Score 28.6;  DB 6;  Length 921;
  Best Local Similarity   47.1%;  Pred. No. 5.5;
  Matches   88;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||  | |  |  | || | ||  | | | | |      |  || |    | ||
Db        417 AACGATGCCATTCAGACCAATACCAACAGCAACGAAAACTCAAGTCCAAATGAGAACAAT 476

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |  ||||   |  ||||    | |   |  | ||   |||| |    ||    || |
Db        477 GCCGTTCAGGCCTCCAATAACGAAAATGCAGGTCAGAACGAGAACGAGAACAATGCTACT 536

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || | || || |  | |||| |  | | || |  |      |||   ||    |||
Db        537 CACAGCTTCAACGACAGCAACGGGAGCAGCAATTACGACACAAATACAGATGGAAACACC 596

Qy        361 CATGAAA 367
              ||  | |
Db        597 CAGAATA 603


RESULT 13
PCT-US04-05654-1030
; Sequence 1030, Application PC/TUS0405654
; GENERAL INFORMATION:
;  APPLICANT: Sherman, Bradley K
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline  E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Dubell III, Arnold N
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Repetti, Peter
;  APPLICANT:  Century, Karen
;  APPLICANT:  Gutterson, Neal
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Kumimoto, Roderick W
;  APPLICANT:  Pilgrim, Marsha L
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 PCT
;  CURRENT APPLICATION NUMBER: PCT/US04/05654
;  CURRENT FILING DATE:  2004-03-03
;  PRIOR APPLICATION NUMBER: 10/374,780
;  PRIOR FILING DATE: 2003-02-25
;  PRIOR APPLICATION NUMBER: 10/675,852
;  PRIOR FILING DATE: 2003-09-30
;  NUMBER OF SEQ ID NOS: 2950
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1030
;   LENGTH: 1186
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
PCT-US04-05654-1030

  Query Match             6.3%;  Score 28.6;  DB 1;  Length 1186;
  Best Local Similarity   59.0%;  Pred. No. 6.1;
  Matches   49;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         44 CTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCG 103
```

```
Db        606 CAGGGACTGCTCCAGCGGCCGGCGGCGGCTACGGCGGTGGCGGCGGCGGTGGTGGTGGCG 665

Qy        104 GCGGCAATAGTTCCGGCCCGGAC 126
              || || | |  | ||||| || |
Db        666 GCTGCTACAACTGCGGCCAGGCC 688


RESULT 14
US-10-767-701-5378
; Sequence 5378, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 5378
;   LENGTH: 596
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS127272_1
US-10-767-701-5378

  Query Match             6.2%;  Score 28.2;  DB 6;  Length 596;
  Best Local Similarity   53.1%;  Pred. No. 6.1;
  Matches   60;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy         58 GCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCC 117
              || | || ||      |   |  |||||||||  |  | | |||||| |||  ||   |
Db        105 GCCGCCGCCGCCGGCGACGTGAACGGCGGCGGCTACTACACCGGCGGCAGCAGCAGCAGC 164

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCT 170
              | |   | | ||   |   |     ||  ||||||  ||| | | | ||  ||
Db        165 GCCGGCGGCCCATACTCCGGTTACAATGCGTACGGCGCCGGTGAGGGTGGCCT 217


RESULT 15
US-10-767-701-8781/c
; Sequence 8781, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 8781
;   LENGTH: 610
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS55918_1
US-10-767-701-8781

  Query Match             6.2%;  Score 28.2;  DB 6;  Length 610;
  Best Local Similarity   55.7%;  Pred. No. 6.2;
  Matches   54;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy         26 TCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCG 85
              || |||||| ||   || | | |||  | ||   ||| | ||       |  || ||||||
Db        119 TCACAGCAAACGAGTTTCCAGCCAGCCCGCTCCGTGGTGCCGAGTTGGAGCGGCGCGGCG 60

Qy         86 GCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCC 122
              ||||    |      |||||| |    | | |||| |

Db         59 GCGGCGGACTCGGGTGCGGCGACGGGCGCTTCGGCGC 23


Search completed: March 15, 2004, 13:47:24
Job time : 54.4 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 02:55:53 ; Search time 1963.4 Seconds
(without alignments)
6935.497 Million cell updates/sec

Title: US-09-543-407-23
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	239	52.4	680	12	BJ618688	BJ618688 BJ618688
c	2	37.6	8.2	917	13	BQ689172	BQ689172 AGENCOURT
	3	37	8.1	415	12	BG984597	BG984597 CM0-CN016
c	4	37	8.1	434	14	T78135	T78135 yc98c10.r1
c	5	37	8.1	514	14	CF139347	CF139347 UI-HF-CB0
c	6	37	8.1	522	12	BI908548	BI908548 603069939
c	7	37	8.1	593	14	CF131618	CF131618 UI-HF-FQ0
c	8	37	8.1	638	10	BE274107	BE274107 601120625
c	9	37	8.1	743	12	BI223663	BI223663 602943576
c	10	37	8.1	762	12	BG166754	BG166754 602339283
c	11	37	8.1	788	10	BE797741	BE797741 601582219
c	12	37	8.1	840	12	BG821996	BG821996 602726102
c	13	37	8.1	842	13	BU192328	BU192328 AGENCOURT
c	14	37	8.1	848	13	BX458187	BX458187 BX458187
c	15	37	8.1	855	13	BU176383	BU176383 AGENCOURT
c	16	37	8.1	866	13	BU501035	BU501035 AGENCOURT
c	17	37	8.1	870	9	AL529046	AL529046 AL529046
c	18	37	8.1	891	13	BQ221454	BQ221454 AGENCOURT
c	19	37	8.1	891	13	BU181949	BU181949 AGENCOURT
c	20	37	8.1	892	13	BQ653899	BQ653899 AGENCOURT
c	21	37	8.1	893	14	CD514585	CD514585 AGENCOURT
c	22	37	8.1	898	13	BQ948515	BQ948515 AGENCOURT
c	23	37	8.1	899	13	BU553911	BU553911 AGENCOURT
c	24	37	8.1	906	13	BQ433516	BQ433516 AGENCOURT
c	25	37	8.1	907	13	BQ425145	BQ425145 AGENCOURT
c	26	37	8.1	907	13	BU157531	BU157531 AGENCOURT
c	27	37	8.1	908	13	BQ961790	BQ961790 AGENCOURT
c	28	37	8.1	915	13	BQ068188	BQ068188 AGENCOURT
c	29	37	8.1	916	13	BQ920773	BQ920773 AGENCOURT
c	30	37	8.1	918	13	BQ644445	BQ644445 AGENCOURT
c	31	37	8.1	929	13	BU500508	BU500508 AGENCOURT
c	32	37	8.1	938	13	BQ642938	BQ642938 AGENCOURT
c	33	37	8.1	945	10	BF036891	BF036891 601460390
c	34	37	8.1	947	13	BU931099	BU931099 AGENCOURT
c	35	37	8.1	963	9	AL560441	AL560441 AL560441
c	36	37	8.1	972	9	AL555785	AL555785 AL555785
c	37	37	8.1	973	13	BQ064771	BQ064771 AGENCOURT
c	38	37	8.1	1008	12	BM923895	BM923895 AGENCOURT
c	39	37	8.1	1049	9	AL560700	AL560700 AL560700
c	40	37	8.1	1088	12	BM557923	BM557923 AGENCOURT
c	41	37	8.1	1099	9	AL558356	AL558356 AL558356
c	42	37	8.1	1177	9	AL524323	AL524323 AL524323
c	43	37	8.1	1201	9	AL583329	AL583329 AL583329
c	44	37	8.1	2028	29	AY404937	AY404937 Homo sapi
c	45	37	8.1	2028	29	AY404938	AY404938 Pan trogl

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

```
          http://xenopus.nibb.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match             52.4%;  Score 239;  DB 12;  Length 680;
  Best Local Similarity   70.2%;  Pred. No. 2.2e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||   | ||    | ||| |||||| |||||||||   || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||   ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
              || ||  | || ||||||||||  ||||| | |||   |  |||||  |   | | |
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||     |   | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138


RESULT 2
BQ689172/c
LOCUS       BQ689172                 917 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8047272 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6207593
            5', mRNA sequence.
ACCESSION   BQ689172
VERSION     BQ689172.1  GI:21814488
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 917)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2363  row: o  column: 18
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .917
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6207593"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of  California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             8.2%;  Score 37.6;  DB 13;  Length 917;
  Best Local Similarity   51.8%;  Pred. No. 12;
  Matches   85;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        284 ATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGG 343
              | ||   |||| ||  ||||||  |||||      | ||||  |  ||   |||     | |
Db        695 AAGGCCACAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTG 636

Qy        344 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCG 403
                | || |||| |  |||  |   |   | |     || | || | |   ||||  |   |
Db        635 GAAACCATGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGG 576

Qy        404 TAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              |   | | | ||||  ||| | |  || || |||      || ||
Db        575 TCTGGCTTCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 532


RESULT 3
BG984597
LOCUS       BG984597                 415 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  CM0-CN0165-060401-809-e04 CN0165 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG984597
VERSION     BG984597.1  GI:14387332
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
```

(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-CN0165-060401-809-e04&t3=2001-04-06&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 414.

FEATURES Location/Qualifiers
source 1. .415
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CN0165"
/note="Organ: colon_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 8.1%; Score 37; DB 12; Length 415;
Best Local Similarity 52.2%; Pred. No. 12;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        125 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 184

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |  ||||  |   ||   | |
Db        185 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 244

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  ||| | |  || || |||     || ||
Db        245 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 281
```

RESULT 4
T78135/c
LOCUS T78135 434 bp mRNA linear EST 07-MAR-1995
DEFINITION yc98c10.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:24221 5' similar to gb:J03068 ACYLAMINO-ACID-RELEASING ENZYME (HUMAN);, mRNA sequence.
ACCESSION T78135
VERSION T78135.1 GI:696644
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 434)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 2012
High quality sequence stops: 323 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2012 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 323.

FEATURES Location/Qualifiers
source 1. .434
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:396568"
/db_xref="taxon:9606"
/clone="IMAGE:24221"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares infant brain 1NIB"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not I; Site_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 434;
Best Local Similarity 52.2%; Pred. No. 13;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        267 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 208

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |  ||||  |   ||   | |
Db        207 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 148

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  ||| | |  || || |||     || ||
Db        147 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 111
```

RESULT 5
CF139347/c
LOCUS CF139347 514 bp mRNA linear EST 06-AUG-2003
DEFINITION UI-HF-CB0-aow-g-08-0-UI.r1 NIH_MGC_210 Homo sapiens cDNA clone IMAGE:3096759 5', mRNA sequence.
ACCESSION CF139347
VERSION CF139347.1 GI:33254791
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 514)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Tim Ratlift
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at http://genome.uiowa.edu/distribution/humanfl.html
Seq primer: pYX-5.

FEATURES Location/Qualifiers
source 1. .514

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3096759"
/tissue_type="CNCAP(3)T-225 cell line"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_MGC_210"
/note="Organ: Prostate; Vector: pT7T3 Pac; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pT7T3 Pac vector. The library tag sequence located between the Not I site and the polyA tail is CCCAC. Tissue was provided by Tim Ratlift."

ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 514;
Best Local Similarity 52.2%; Pred. No. 14;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||      | ||||  |  ||   |||     | |  | ||
Db        214 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 155

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |   ||||  |   ||  | |
Db        154 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 95

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  |||  | |  || || |||     || ||
Db         94 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 58
```

RESULT 6
BI908548/c
LOCUS BI908548 522 bp mRNA linear EST 16-OCT-2001
DEFINITION 603069939F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5219120 5', mRNA sequence.
ACCESSION BI908548
VERSION BI908548.1 GI:16171539
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 522)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM11550 row: m column: 09
High quality sequence stop: 522.
FEATURES Location/Qualifiers
source 1. .522
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5219120"
/tissue_type="leukocyte"
/lab_host="DH10B"
/clone_lib="NIH_MGC_118"
/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 8.1%; Score 37; DB 12; Length 522;
Best Local Similarity 52.2%; Pred. No. 14;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||      | ||||  |  ||   |||     | |  | ||
Db        453 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 394

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |   ||||  |   ||  | |
Db        393 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 334

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  |||  | |  || || |||     || ||
Db        333 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 297
```

RESULT 7
CF131618/c
LOCUS CF131618 593 bp mRNA linear EST 05-AUG-2003
DEFINITION UI-HF-FQ0-awr-k-15-0-UI.r1 NIH_MGC_215 Homo sapiens cDNA clone IMAGE:30555422 5', mRNA sequence.
ACCESSION CF131618
VERSION CF131618.1 GI:33214055
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 593)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Mary Hendrix
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at http://genome.uiowa.edu/distribution/humanfl.html
Seq primer: pYX-5.
FEATURES Location/Qualifiers
source 1. .593
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30555422"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_MGC_215"
/note="Organ: Lung; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAAGGCCA. Tissue was provided by Mary Hendrix."

ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 593;
Best Local Similarity 52.2%; Pred. No. 14;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        183 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 124

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |   ||||  |   ||   | |
Db        123 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 64

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  ||| | |  || || |||     || ||
Db         63 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 27
```

RESULT 8
BE274107/c
LOCUS BE274107 638 bp mRNA linear EST 13-JUL-2000
DEFINITION 601120625F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2967078 5', mRNA sequence.
ACCESSION BE274107
VERSION BE274107.1 GI:9149046
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 638)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM71 row: b column: 07
High quality sequence stop: 636.
FEATURES Location/Qualifiers
source 1. .638
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2967078"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_20"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match 8.1%; Score 37; DB 10; Length 638;
Best Local Similarity 52.2%; Pred. No. 15;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        471 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 412

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |   ||||  |   ||   | |
Db        411 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 352

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  ||| | |  || || |||     || ||
Db        351 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 315
```

RESULT 9
BI223663/c
LOCUS BI223663 743 bp mRNA linear EST 11-JUL-2001
DEFINITION 602943576F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5106499 5', mRNA sequence.
ACCESSION BI223663
VERSION BI223663.1 GI:14677107
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 743)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM11257 row: h column: 20
High quality sequence stop: 739.
FEATURES Location/Qualifiers
source 1. .743
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5106499"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_lib="NIH_MGC_12"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.4 kb. Library prepared by Life Technologies."

ORIGIN

Query Match 8.1%; Score 37; DB 12; Length 743;
Best Local Similarity 52.2%; Pred. No. 16;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        706 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 647

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |   ||||  |   ||   | |
Db        646 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 587

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  ||| | |  || || |||     || ||
Db        586 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 550
```

```
RESULT 10
BG166754/c
LOCUS       BG166754                762 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602339283F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4447125 5',
            mRNA sequence.
ACCESSION   BG166754
VERSION     BG166754.1  GI:12673457
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 762)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10227  row: f  column: 22
            High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1. .762
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4447125"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_89"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             8.1%;  Score 37;  DB 12;  Length 762;
  Best Local Similarity   52.2%;  Pred. No. 16;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        576 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 517

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |   || | || | |   ||||  |   ||   | |
Db        516 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 457

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
                | ||||  ||| | |  || || |||     || ||
Db        456 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 420


RESULT 11
BE797741/c
LOCUS       BE797741                788 bp    mRNA    linear   EST 20-SEP-2000
DEFINITION  601582219F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3936748 5',
            mRNA sequence.
ACCESSION   BE797741
VERSION     BE797741.1  GI:10218939
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM781  row: e  column: 05
            High quality sequence start: 3
            High quality sequence stop: 783.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3936748"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_7"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

  Query Match             8.1%;  Score 37;  DB 10;  Length 788;
  Best Local Similarity   52.2%;  Pred. No. 16;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        554 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 495

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |   || | || | |   ||||  |   ||   | |
Db        494 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 435

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
                | ||||  ||| | |  || || |||     || ||
Db        434 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 398


RESULT 12
BG821996/c
LOCUS       BG821996                840 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  602726102F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4865593 5',
            mRNA sequence.
ACCESSION   BG821996
VERSION     BG821996.1  GI:14169583
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 840)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
```

```
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1731  row: c  column: 02
           High quality sequence stop: 745.
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4865593"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_15"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
ORIGIN

  Query Match             8.1%;  Score 37;  DB 12;  Length 840;
  Best Local Similarity   52.2%;  Pred. No. 17;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        535 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 476

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |  ||||  |   ||   | |
Db        475 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 416

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
                | ||||  |||  | |  || || |||     || ||
Db        415 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 379


RESULT 13
BU192328/c
LOCUS       BU192328                842 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7937874 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6013184
            5', mRNA sequence.
ACCESSION   BU192328
VERSION     BU192328.1  GI:22706303
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 842)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM13205  row: k  column: 09
           High quality sequence stop: 604.
FEATURES             Location/Qualifiers
     source          1. .842
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6013184"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_92"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             8.1%;  Score 37;  DB 13;  Length 842;
  Best Local Similarity   52.2%;  Pred. No. 17;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        531 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 472

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |  ||||  |   ||   | |
Db        471 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 412

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | ||||  |||  | |  || || |||     || ||
Db        411 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 375


RESULT 14
BX458187/c
LOCUS       BX458187                848 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX458187 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CS0DA011YH16 5-PRIME, mRNA sequence.
ACCESSION   BX458187
VERSION     BX458187.1  GI:31021002
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 848)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2188.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DA011DD08QP1&cluster=2188.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DA011DD08QP1.
FEATURES             Location/Qualifiers
     source          1. .848
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA011YH16"
                     /tissue_type="NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN
```

```
  Query Match             8.1%;  Score 37;  DB 13;  Length 848;
  Best Local Similarity   52.2%;  Pred. No. 17;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        788 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 729

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |  |   | |    || | || | |   ||||  |   ||   | |
Db        728 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 669

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
               | ||||  ||| | |  || || |||     || ||
Db        668 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 632


RESULT 15
BU176383/c
LOCUS       BU176383                855 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7909215 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6092741
            5', mRNA sequence.
ACCESSION   BU176383
VERSION     BU176383.1  GI:22690367
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 855)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13360  row: n  column: 06
            High quality sequence stop: 653.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6092741"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_70"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
ORIGIN

  Query Match             8.1%;  Score 37;  DB 13;  Length 855;
  Best Local Similarity   52.2%;  Pred. No. 17;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCG 350
              |||| ||  ||||||  |||||     | ||||  |  ||   |||    | |  | ||
Db        491 CAGAGATGTTGCCACTCTCGACATCCAGCACGCAGAGCACAGGGATGCTTTTGGAAACCA 432

Qy        351 TGTTGTTACCCATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGT 410
              ||||  |  |||  |   |   | |    || | || | |   ||||  |   ||   | |
Db        431 TGTTTTCTCCCCAGTCTTCATAAAACACAAACTGATCCCCCTTGATGGCTTGGTCTGGCT 372

Qy        411 GCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
               | ||||  ||| | |  || || |||     || ||
Db        371 TCTTCAGCCTGGCTATCTCATCATCGCTGGCACTGAC 335


Search completed: March 15, 2004, 10:09:24
Job time : 1967.4 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 19:02:18 ; Search time 2269.8 Seconds
(without alignments)
8707.562 Million cell updates/sec

Title: US-09-543-407-25
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	404.8	88.8	456	6	I44909	I44909 Sequence 58
2	404.8	88.8	2067	1	SEU43280	U43280 Salmonella
3	396.8	87.0	5103	1	STAJ2301	AJ002301 Salmonell
4	396.8	87.0	22411	1	AE008749	AE008749 Salmonell
5	395.2	86.7	254050	1	AL627269	AL627269 Salmonell
c 6	395.2	86.7	301983	1	AE016840	AE016840 Salmonell
7	389.4	85.4	1048	1	STAGFBA	AJ000514 Salmonell
8	280.6	61.5	361	6	I44908	I44908 Sequence 56
9	267.2	58.6	2889	1	CSP515700	AJ515700 Citrobact
10	229.2	50.3	2920	1	CFR515701	AJ515701 Citrobact
11	222.4	48.8	4680	1	ECCSGABDG	X90754 E.coli csgG
12	222.4	48.8	10346	1	AE000205	AE000205 Escherich
13	222.4	48.8	15047	1	D90741	D90741 Escherichia
14	220.8	48.4	456	6	AX814811	AX814811 Sequence
15	220.8	48.4	648	1	ECOCSGAA	L04979 Escherichia
16	219	48.0	1711	1	AF275733	AF275733 Escherich
17	219	48.0	10190	1	AE005315	AE005315 Escherich
18	219	48.0	327773	1	AP002554	AP002554 Escherich
19	215.8	47.3	306358	1	AE016759	AE016759 Escherich
20	192	42.1	2883	1	ESA515702	AJ515702 Enterobac
21	177.8	39.0	230	1	SEU53207	U53207 Salmonella
22	156.8	34.4	10370	1	AE015131	AE015131 Shigella
23	156.8	34.4	292504	1	AE016981	AE016981 Shigella
c 24	133.8	29.3	437	1	AF237726	AF237726 Shigella
25	93.2	20.4	19201	1	D90742	D90742 Escherichia
26	64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
c 27	49.6	10.9	2000	6	AX655393	AX655393 Sequence
28	48.2	10.6	78	6	AX814809	AX814809 Sequence
29	39	8.6	5772	3	AF466309	AF466309 Dictyoste
30	37.8	8.3	233764	2	AC106430	AC106430 Rattus no
31	37.4	8.2	2678	10	BC011097	BC011097 Mus muscu
c 32	37.4	8.2	165074	9	AC093519	AC093519 Homo sapi
c 33	37.4	8.2	217726	10	AL928680	AL928680 Mouse DNA
c 34	36.6	8.0	1554	6	AX654234	AX654234 Sequence
c 35	36.6	8.0	2075	8	AK069542	AK069542 Oryza sat
c 36	36.6	8.0	116519	2	AP003992	AP003992 Oryza sat
37	36.4	8.0	87967	8	AC005223	AC005223 Arabidops
c 38	36.4	8.0	153108	9	AL355273	AL355273 Human DNA
39	36.2	7.9	78631	9	HSAC000057	AC000057 Human BAC
c 40	36.2	7.9	186380	9	AC100822	AC100822 Homo sapi
41	36	7.9	1168	6	AX647633	AX647633 Sequence
c 42	36	7.9	2428	6	AX833196	AX833196 Sequence
c 43	36	7.9	2428	9	AK094755	AK094755 Homo sapi
c 44	36	7.9	134369	9	AC104825	AC104825 Homo sapi
45	36	7.9	153756	9	AC073866	AC073866 Homo sapi

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match              88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity    93.0%;  Pred. No. 1.4e-101;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match              88.8%;  Score 404.8;  DB 1;  Length 2067;
  Best Local Similarity    93.0%;  Pred. No. 1.5e-101;
```

```
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||   |    ||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||      |        ||  | ||    ||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match             87.0%;  Score 396.8;  DB 1;  Length 5103;
  Best Local Similarity   91.9%;  Pred. No. 2.5e-99;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES Location/Qualifiers

source 1. .22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

gene 434. .1308
/gene="phoH"
/note="synonym: STM1126"

RBS 434. .439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"

CDS 454. .1308
/gene="phoH"
/note="similar to E. coli PhoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=11
/product="PhoB-dependent ATP-binding pho regulon component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
QRTLHAYS"

gene complement(1414. .2303)
/gene="STM1127"

CDS complement(1414. .2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29% identity in aa 15 - 264"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
QARESKLRTMNALDMTK"

RBS complement(2298. .2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"

gene complement(2580. .4088)
/gene="STM1128"

CDS complement(2580. .4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22% identity in aa 7 - 478"
/codon_start=1
/transl_table=11
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
DKFFPPATQPHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"

RBS complement(4083. .4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"

gene complement(4413. .5099)
/gene="STM1129"

CDS complement(4413. .5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 - 226"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
WFSHAVKR"

RBS complement(5094. .5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"

gene 5599. .6759
/gene="STM1130"

CDS 5599. .6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41% identity in aa 40 - 403"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"

gene 6793. .7497
/gene="STM1131"

RBS 6793. .6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           87.0%;  Score 396.8;  DB 1;  Length 22411;
  Best Local Similarity   91.9%;  Pred. No. 2.6e-99;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||| || |||||||||||||||||||||||||||||||||||||||  |   ||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSPEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSPSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             86.7%;  Score 395.2;  DB 1;  Length 254050;
  Best Local Similarity   91.7%;  Pred. No. 7.7e-99;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||   |   ||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.7%;  Score 395.2;  DB 1;  Length 301983;
Best Local Similarity   91.7%;  Pred. No. 7.8e-99;

  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||  |   ||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             85.4%;  Score 389.4;  DB 1;  Length 1048;
  Best Local Similarity   91.0%;  Pred. No. 2.8e-97;
  Matches  414;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||   |    ||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||      |          ||  | ||    ||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  ||||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             61.5%;  Score 280.6;  DB 6;  Length 361;
  Best Local Similarity   89.9%;  Pred. No. 5.2e-67;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTG 183
              |||||||||||||||||||||||||||||||||||||||||||||||||   |    ||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTA 243
              | | |||||      |          ||  | ||    ||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           58.6%;  Score 267.2;  DB 1;  Length 2889;
  Best Local Similarity  75.7%;  Pred. No. 2.9e-63;
  Matches  345;  Conservative    0;  Mismatches  108;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |    ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||| || ||  |||||||||||||||  ||   |  ||||||  ||||||      ||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                |   | |||     |         ||    ||    || | ||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || ||||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               |||||||||| ||||| || ||||| || || ||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
CFR515701
LOCUS       CFR515701              2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
  Query Match             50.3%;  Score 229.2;  DB 1;  Length 2920;
  Best Local Similarity   70.6%;  Pred. No. 1.1e-52;
  Matches  322;  Conservative    0;  Mismatches  128;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||       |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              || |||||  |  |||| || || || || || ||    ||| ||||   |  |   ||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | | |||              ||  ||||    ||   ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||||| ||| | ||||| |||    ||||  ||||  || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| ||||| |||||||| || |||||  | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||| |||||||||||||| || ||||| ||||  ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || |||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

```
RESULT 11
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
```

```
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           48.8%;  Score 222.4;  DB 1;  Length 4680;
  Best Local Similarity   68.0%;  Pred. No. 8.7e-51;
  Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||   | ||   | ||| |||||| |||||||||  || |||| |||| |||  |   ||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |  |       ||  | ||    ||    |||||  || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 12
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
```

```
          and products; all new functional assignments courtesy of Monica
          Riley; added promoters, protein binding sites, and repeated
          sequences described in reference 1. The unique numeric identifiers
          beginning with a lowercase 'b' assigned to each gene (protein- or
          RNA-encoding) are now designated as gene synonyms instead of
          labels. This should allow them to be searched for in Entrez as gene
          names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

Query Match             48.8%;  Score 222.4;  DB 1;  Length 10346;
Best Local Similarity   68.0%;  Pred. No. 8.9e-51;
Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| |||  |   ||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              . | | |||||     |  |       ||  | ||     ||     ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

RESULT 13
D90741
LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
FEATURES Location/Qualifiers
source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER

```
                ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS             2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
CDS             3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
gene            complement(4117. .4950)
                /gene="csgG"
CDS             complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
gene            complement(4977. .5393)
                /gene="csgF"
CDS             complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene            complement(5418. .5807)
                /gene="csgE"
CDS             complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
gene            complement(5812. .6462)
                /gene="csgD"
CDS             complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene            7217. .7672
                /gene="csgB"
CDS             7217. .7672
```

Query Match 48.8%; Score 222.4; DB 1; Length 15047;
Best Local Similarity 68.0%; Pred. No. 9e-51;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| |||  |   ||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |  |       ||  | ||    ||    ||||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||  ||||| || || || || ||    || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168
```

```
RESULT 14
AX814811
LOCUS       AX814811                456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
```

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match             48.4%;  Score 220.8;  DB 6;  Length 456;
  Best Local Similarity   67.8%;  Pred. No. 2.3e-50;
  Matches  309;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| |||  |   ||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |  |       ||  | ||     ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match             48.4%;  Score 220.8;  DB 1;  Length 648;
  Best Local Similarity   67.8%;  Pred. No. 2.3e-50;
  Matches  309;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| |||  |   ||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |  |       ||  | ||     ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 15, 2004, 07:24:11
Job time : 2274.2 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 18:52:53 ; Search time 290.4 Seconds
(without alignments)
6670.724 Million cell updates/sec

Title: US-09-543-407-25
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64629	Aac64629 AgfA::PT3
	2	404.8	88.8	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	404.8	88.8	456	2	AAT74142	Aat74142 Salmonell
	4	404.8	88.8	456	3	AAC64617	Aac64617 Salmonell
	5	385.6	84.6	456	3	AAC64626	Aac64626 AgfA::PT3
	6	355.2	77.9	456	3	AAC64625	Aac64625 AgfA::PT3
	7	353.6	77.5	456	3	AAC64628	Aac64628 AgfA::PT3
	8	353.6	77.5	456	3	AAC64622	Aac64622 AgfA::PT3
	9	353.6	77.5	456	3	AAC64623	Aac64623 AgfA::PT3
	10	352	77.2	456	3	AAC64624	Aac64624 AgfA::PT3
	11	350.4	76.8	456	3	AAC64630	Aac64630 AgfA::PT3
	12	350.4	76.8	456	3	AAC64627	Aac64627 AgfA::PT3
	13	347.2	76.1	456	3	AAC64631	Aac64631 AgfA::PT3
	14	280.6	61.5	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	280.6	61.5	361	2	AAT74141	Aat74141 Salmonell
	16	220.8	48.4	456	3	AAC64619	Aac64619 Escherich
	17	220.8	48.4	456	9	ACF36153	Acf36153 E. coli C
	18	208.4	45.7	646	2	AAQ62647	Aaq62647 Fibronect
	19	146.4	32.1	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
c	21	50.4	11.1	78	3	AAC64609	Aac64609 AgfA (SEF
c	22	49.6	10.9	2000	7	ADA71938	Ada71938 Rice gene
	23	48.2	10.6	78	3	AAC64610	Aac64610 AgfA (SEF
	24	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	25	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	26	43.4	9.5	78	3	AAC64606	Aac64606 SefA (SEF
	27	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	28	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
c	29	40.6	8.9	78	3	AAC64605	Aac64605 SefA (SEF
	30	40	8.8	100	7	ACD68809	Acd68809 E. coli K
	31	36.6	8.0	456	3	AAC64620	Aac64620 Escherich
c	32	36.6	8.0	1554	7	ADA70781	Ada70781 Rice gene
	33	36	7.9	1168	9	ADC87372	Adc87372 Human GPC
	34	35.8	7.9	72	9	ACF36148	Acf36148 E. coli C
	35	34.2	7.5	456	3	AAC64618	Aac64618 Salmonell
	36	34.2	7.5	1050	3	AAF13281	Aaf13281 Aspergill
	37	34.2	7.5	2000	7	ADA71486	Ada71486 Rice gene
	38	33.8	7.4	2547	7	ACA48284	Aca48284 Prokaryot
	39	33.4	7.3	1355	7	ABZ37481	Abz37481 M13-III n
c	40	33.2	7.3	392	7	ACA37361	Aca37361 Prokaryot
	41	32.8	7.2	1083	5	AAS76745	Aas76745 DNA encod
	42	32.8	7.2	3562	4	ABL06749	Abl06749 Drosophil
	43	32.6	7.1	2553	3	AAA09500	Aaa09500 V. furnis
c	44	32.4	7.1	901	4	ABL09481	Abl09481 Drosophil
c	45	32.4	7.1	2663	4	ABL24658	Abl24658 Drosophil

ALIGNMENTS

RESULT 1
AAC64629
ID AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.2e-139;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 2; Length 456;
Best Local Similarity 93.0%; Pred. No. 4.9e-122;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 2;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 4.9e-122;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |         ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
Best Local Similarity 93.0%; Pred. No. 4.9e-122;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||   |    ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 84.6%; Score 385.6; DB 3; Length 456;
Best Local Similarity 90.4%; Pred. No. 9.5e-116;
Matches 412; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||   |    ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 8.7e-106;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              |||||||       |    ||   |      | | ||  | |    || | |       ||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | ||||       |          ||  | ||     ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 2.9e-105;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||    ||   ||    |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 8
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 2.9e-105;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||| |         | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match             77.5%;  Score 353.6;  DB 3;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 2.9e-105;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64624
ID   AAC64624 standard; DNA; 456 BP.
XX
AC   AAC64624;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 9.7e-105;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||  ||    | | | |   | |     |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 3.2e-104;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |||   |       | |  ||   |    |       |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 3.2e-104;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||  |  |       | ||  |    ||  |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             76.1%;  Score 347.2;  DB 3;  Length 456;
  Best Local Similarity   85.1%;  Pred. No. 3.6e-103;
  Matches  388;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||  |     ||  | ||   ||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |   |                |   | |  ||   ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             61.5%;  Score 280.6;  DB 2;  Length 361;
  Best Local Similarity   89.9%;  Pred. No. 2.1e-81;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTG 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTA 243
              | | |||||     |          ||  | ||    ||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             61.5%;  Score 280.6;  DB 2;  Length 361;
  Best Local Similarity   89.9%;  Pred. No. 2.1e-81;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTG 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTA 243
              | | |||||     |          ||  | ||    ||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 15, 2004, 04:14:43
```

Job time : 292.4 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:01:03 ; Search time 56.8 Seconds
(without alignments)
4455.240 Million cell updates/sec

Title: US-09-543-407-25
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	404.8	88.8	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	280.6	61.5	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	33	7.2	1038	4	US-09-543-681A-2664	Sequence 2664, Ap
c	4	32	7.0	1344	3	US-09-120-927-1	Sequence 1, Appli
c	5	32	7.0	1344	4	US-09-431-614-5	Sequence 5, Appli
c	6	32	7.0	1647	4	US-09-252-991A-9113	Sequence 9113, Ap
	7	32	7.0	2001	4	US-09-252-991A-8873	Sequence 8873, Ap
	8	32	7.0	14770	4	US-09-220-132-30	Sequence 30, Appl
c	9	31.6	6.9	1338	4	US-09-252-991A-7783	Sequence 7783, Ap
	10	31.6	6.9	1662	1	US-08-565-386-2	Sequence 2, Appli
	11	31.6	6.9	1905	4	US-09-252-991A-7560	Sequence 7560, Ap
	12	31.6	6.9	4529	1	US-08-565-386-1	Sequence 1, Appli
c	13	31.4	6.9	275	4	US-09-313-294A-5838	Sequence 5838, Ap
	14	31.4	6.9	1530	4	US-09-543-681A-1772	Sequence 1772, Ap
	15	31.4	6.9	2355	4	US-08-913-159-12	Sequence 12, Appl
	16	31.4	6.9	4061	4	US-09-132-769-4	Sequence 4, Appli
	17	31.4	6.9	51259	3	US-08-781-891-209	Sequence 209, App
	18	31.4	6.9	51259	4	US-09-618-166-209	Sequence 209, App
	19	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
	20	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	21	30.6	6.7	2718	3	US-09-074-658-69	Sequence 69, Appl
c	22	30.6	6.7	11707	3	US-09-136-574A-1	Sequence 1, Appli
	23	30.6	6.7	49617	4	US-09-596-002-28	Sequence 28, Appl
c	24	30.2	6.6	999	4	US-09-252-991A-14326	Sequence 14326, A
	25	30.2	6.6	1251	4	US-09-252-991A-14352	Sequence 14352, A
	26	30.2	6.6	1593	4	US-09-252-991A-14376	Sequence 14376, A
c	27	30.2	6.6	1626	4	US-09-252-991A-14296	Sequence 14296, A
c	28	30.2	6.6	3636	4	US-08-961-527-78	Sequence 78, Appl
	29	30.2	6.6	9578	4	US-08-961-527-127	Sequence 127, App
c	30	30	6.6	1278	4	US-09-252-991A-738	Sequence 738, App
c	31	30	6.6	1359	4	US-09-252-991A-705	Sequence 705, App
	32	30	6.6	1587	4	US-09-252-991A-676	Sequence 676, App
	33	30	6.6	3573	3	US-09-353-585-4	Sequence 4, Appli
	34	30	6.6	6471	3	US-09-353-585-1	Sequence 1, Appli
	35	29.8	6.5	357	4	US-09-489-039A-1118	Sequence 1118, Ap
c	36	29.8	6.5	816	4	US-09-489-039A-1109	Sequence 1109, Ap
	37	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	38	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
c	39	29.8	6.5	1323	4	US-09-252-991A-5147	Sequence 5147, Ap
c	40	29.8	6.5	1422	4	US-09-252-991A-5119	Sequence 5119, Ap
	41	29.8	6.5	1467	4	US-09-252-991A-5070	Sequence 5070, Ap
	42	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
	43	29.8	6.5	4919	3	US-08-456-200B-2	Sequence 2, Appli
	44	29.6	6.5	483	4	US-09-540-236-594	Sequence 594, App
c	45	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.3e-130;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |      ||  | ||    ||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             61.5%;  Score 280.6;  DB 1;  Length 361;
  Best Local Similarity   89.9%;  Pred. No. 1.5e-87;
  Matches  301;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTG 183
              |||||||||||||||||||||||||||||||||||||||||||||||||   |   ||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTA 243
              | | |||||     |      ||  | ||    ||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

```
RESULT 3
US-09-543-681A-2664/c
; Sequence 2664, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2664
;   LENGTH: 1038
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-2664

  Query Match             7.2%;  Score 33;  DB 4;  Length 1038;
  Best Local Similarity   48.6%;  Pred. No. 0.27;
  Matches   90;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        156 CGCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 215
              | | |  | | |  ||||  |||||||||  |||  | | || ||      | |   ||
Db        343 CACCATAGGTACAACTTATCATCAGCTGGCAACCAATCTGGTGACAATGCATCTCTAACT 284

Qy        216 TGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACT 275
              |    | ||  | | |   |  | || || |||      | ||||    || || |   |
Db        283 TTTCCGCTAAAGCAGCAAAGAAGGGGTTGGACAGCCGAGGAATAACCAGACCATAGGTAT 224

Qy        276 GACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              |   | ||  ||   | | | |  ||| |  | | |  | ||||||| |||
Db        223 CTGTTTTATTGAGTTTTAAGCTTCTTGCCGCATGGTTAACGGTATAACCAAAAACTTCGA 164
```

```
Qy        336 TATTA 340
                || |
Db        163 CATAA 159


RESULT 4
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
;    APPLICANT:  Kim, Jihyun Francis
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;    TITLE OF INVENTION:  ERWINIA AMYLOVORA AND ITS USE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,927
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,108
;      FILING DATE:  06-AUG-1977
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1581
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1344 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.0%;  Score 32;  DB 3;  Length 1344;
  Best Local Similarity   54.2%;  Pred. No. 0.69;
  Matches   65;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90


RESULT 5
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.0%;  Score 32;  DB 4;  Length 1344;
  Best Local Similarity   54.2%;  Pred. No. 0.69;
  Matches   65;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90


RESULT 6
US-09-252-991A-9113/c
; Sequence 9113, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 9113
;   LENGTH:  1647
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9113

  Query Match             7.0%;  Score 32;  DB 4;  Length 1647;
  Best Local Similarity   50.7%;  Pred. No. 0.77;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        183 GGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGT 242
              ||  || |  ||| | | |   | || |||| |||   |  | ||| | | ||| ||||
Db        165 GGCGACGCTCGTTTTCAGCAGAGGAACGGCATATGATCGACAGGGTGAAGCCGCTGATGG 106

Qy        243 AGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGC 302
              |||  |||| | | |  | | |    |     |    |||         | | || | |||
Db        105 CGGCATGGGTTCTGGTCTTCGCATCTTCTTCGTTCGCCAGCGAAACGGCGGGAAAATTGC 46

Qy        303 CACCATCGACCAGTGGAACGCTAAAAACTCCG 334
              |||  | | |   |  | ||  ||||  ||
Db         45 CACGAAAGCCAGCTCTAGGGCCGGAAACCGCG 14


RESULT 7
US-09-252-991A-8873
; Sequence 8873, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
```

```
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 8873
;   LENGTH: 2001
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8873

  Query Match             7.0%;  Score 32;  DB 4;  Length 2001;
  Best Local Similarity   50.7%;  Pred. No. 0.86;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        183 GGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGT 242
              ||  || |  ||| | | |   | || |||| |||   |  | ||| | | ||| ||||
Db        764 GGCGACGCTCGTTTTCAGCAGAGGAACGGCATATGATCGACAGGGTGAAGCCGCTGATGG 823

Qy        243 AGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGC 302
              |||  |||| | | |  | | |     |    |||         | | || | |||
Db        824 CGGCATGGGTTCTGGTCTTCGCATCTTCTTCGTTCGCCAGCGAAACGGCGGGAAAATTGC 883

Qy        303 CACCATCGACCAGTGGAACGCTAAAAACTCCG 334
              ||| |  | |   |  |  ||   ||||  ||
Db        884 CACGAAAGCCAGCTCTAGGGCCGGAAACCGCG 915


RESULT 8
US-09-220-132-30
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  TITLE OF INVENTION:  METHODS AND COMPOSITONS FOR THE IDENTIFICATION AND ASSESSMENT
;  TITLE OF INVENTION:  OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCE
;  FILE REFERENCE:  07334-074001
;  CURRENT APPLICATION NUMBER: US/09/220,132
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  US 60/079,303
;  PRIOR FILING DATE:  1998-03-25
;  PRIOR APPLICATION NUMBER:  US 60/068,821
;  PRIOR FILING DATE:  1997-12-24
;  NUMBER OF SEQ ID NOS:  191
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 30
;   LENGTH: 14770
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-220-132-30

  Query Match             7.0%;  Score 32;  DB 4;  Length 14770;
  Best Local Similarity   48.9%;  Pred. No. 2.7;
  Matches   86;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy        203 ATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 262
              | ||| ||| |  |   || | ||| || |     ||  | ||||  ||    |||
Db      10265 AAGAACTGGAAAGTAATGGATCTGGAAAAGATAATGAATTTGGCCTTGGCCTTGATTCAC 10324

Qy        263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACG 322
              |   | ||||  |  | |||||||||   ||   |  |  ||| ||  ||    |  |  |
Db      10325 CTCAGAATGAAATTGCCCAGAATGGGAACAACGACCAGTCCATCACAGAGTGTTCCATTG 10384

Qy        323 CTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTT 378
              | |  |   | ||  || ||   | |     | | |    | | || | | |||| |
Db      10385 CCACCACAGCAGAGTTTTCTCATGACACGGATGCCACAGAGATCGACTCTCTGGAT 10440


RESULT 9
US-09-252-991A-7783/c
; Sequence 7783, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 7783
;   LENGTH: 1338
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7783

  Query Match             6.9%;  Score 31.6;  DB 4;  Length 1338;
  Best Local Similarity   47.9%;  Pred. No. 0.95;
  Matches   91;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105
              ||||  |||   || ||| ||  | ||  || ||  | ||          |  || ||
Db       1227 GGCATCGCTTCCGCCGGCATCAAGCGACCGGGGCGCAAAGACGTGGTGGTCATGCGCTGC 1168

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165
              | | |  | ||| ||  || |     ||| | | |  |  || | || ||||    |
Db       1167 GCCGAAGGCTCCAGCGTGGCCGGGGTGTTCACCCTGAACGCGTTCTGTGCCGCGCCGGTG 1108

Qy        166 GCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTAT 225
              | ||    | ||||   |  | |||  ||  | |    || |  || | ||  |
Db       1107 ACCCTGGCCAAACAGCGTTTCCTCGGTGAGGTGCGCTACCTGCTGACCAATACCGGCAAC 1048

Qy        226 GGTAACGGCG 235
              |  |||| ||
Db       1047 GCCAACGCCG 1038


RESULT 10
US-08-565-386-2
; Sequence 2, Application US/08565386
; Patent No. 5741697
;  GENERAL INFORMATION:
;    APPLICANT:  Bavoil, Patrik M.
;    APPLICANT:  Hsia, Ru-ching
;    TITLE OF INVENTION:  BACTERIOPHAGE OF CHLAMYDIA PSITTACI
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Roches
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/565,386
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timian, Susan J.
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  176/60040
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1636
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
US-08-565-386-2

  Query Match             6.9%;  Score 31.6;  DB 1;  Length 1662;
  Best Local Similarity   48.9%;  Pred. No. 1.1;
  Matches   85;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        170 TTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTA 229
              |||||   |      ||| || | | | |   || |   ||    ||  ||     | |
Db       1262 TTTATTGGCCTGCTCTTAGCCATTTAGGTGAGCAAGCTGTGCTCAATAAAGAGATTTATT 1321

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              |   |    || ||   ||| |||  | ||| ||| | |||   ||  | || | |  |
Db       1322 GCCAAGGTCCTGCAGTTAAGGATGCTCAGAATGGTAATGTTGTTGTGGATGAGCAAGTCT 1381

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTG 343
              | || || | |   | |   |||     |  ||| |  ||  | |||||||
Db       1382 TTGGATATCAGGAGAGATTTGCGGAGTATCGCTATAAGACTTCGAAAATTACTG 1435


RESULT 11
US-09-252-991A-7560
; Sequence 7560, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 7560
;   LENGTH: 1905
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7560

  Query Match             6.9%;  Score 31.6;  DB 4;  Length 1905;
  Best Local Similarity   47.9%;  Pred. No. 1.2;
  Matches   91;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105
              ||||  |||   || ||| ||   | ||  || ||   | ||       |   || ||
Db        745 GGCATCGCTTCCGCCGGCATCAAGCGACCGGGGCGCAAAGACGTGGTGGTCATGCGCTGC 804

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165
              | | |  | ||| ||  || |     ||| | | |  |   || | || ||||    |
Db        805 GCCGAAGGCTCCAGCGTGGCCGGGGTGTTCACCCTGAACGCGTTCTGTGCCGCGCCGGTG 864

Qy        166 GCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTAT 225
              | ||     | ||||    |   | |||  ||   | ||     || |  || | ||  |
Db        865 ACCCTGGCCAAACAGCGTTTCCTCGGTGAGGTGCGCTACCTGCTGACCAATACCGGCAAC 924

Qy        226 GGTAACGGCG 235
              |  |||| ||
Db        925 GCCAACGCCG 934


RESULT 12
US-08-565-386-1
; Sequence 1, Application US/08565386
; Patent No. 5741697
;  GENERAL INFORMATION:
;    APPLICANT:  Bavoil, Patrik M.
;    APPLICANT:  Hsia, Ru-ching
;    TITLE OF INVENTION:  BACTERIOPHAGE OF CHLAMYDIA PSITTACI
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Roches
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/565,386
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timian, Susan J.
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  176/60040
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1636
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4529 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
US-08-565-386-1

  Query Match             6.9%;  Score 31.6;  DB 1;  Length 4529;
  Best Local Similarity   48.9%;  Pred. No. 1.9;
  Matches   85;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        170 TTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTA 229
              |||||   |      ||| || | | | |   || |   ||    ||  ||     | |
Db       1417 TTTATTGGCCTGCTCTTAGCCATTTAGGTGAGCAAGCTGTGCTCAATAAAGAGATTTATT 1476

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              |   |    || ||   ||| |||  | ||| ||| | |||   ||  | || | |  |
Db       1477 GCCAAGGTCCTGCAGTTAAGGATGCTCAGAATGGTAATGTTGTTGTGGATGAGCAAGTCT 1536

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTG 343
              | || || | |   | |   |||     |  ||| |  ||  | |||||||
Db       1537 TTGGATATCAGGAGAGATTTGCGGAGTATCGCTATAAGACTTCGAAAATTACTG 1590


RESULT 13
US-09-313-294A-5838/c
; Sequence 5838, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 5838
;   LENGTH: 275
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700350760H1
;   NAME/KEY: unsure
;   LOCATION: 60, 65
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5838
```

```
  Query Match             6.9%;  Score 31.4;  DB 4;  Length 275;
  Best Local Similarity   49.1%;  Pred. No. 0.46;
  Matches   83;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        198 TACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGA 257
              | ||||  | ||  ||||||  ||  || |||  |||| | |  ||  ||  |  ||||
Db        274 TCCCCAATCACTGCTACATGCTAGTCCTGATAAATGCGCTGGTTCAGATCAACGCTCGGA 215

Qy        258 TAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              | |  ||   |  || ||  || |      ||||     | ||| || |||      |
Db        214 CGACACGACCGCTCTACAGAACCACACCCAAATCAGGGTACAAGCCGCCGTCGCGAGCAG 155

Qy        318 GAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
                || ||    ||||  |   |  || |  |   |   | | ||||
Db        154 CTGAGCGAAGTGGTCCGCAACGTATCCCGCCGGAAGAGAATGGACTAAC 106


RESULT 14
US-09-543-681A-1772
; Sequence 1772, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1772
;   LENGTH: 1530
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-1772

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 1530;
  Best Local Similarity   59.6%;  Pred. No. 1.2;
  Matches   53;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        185 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAG 244
              || | |    |||||||   ||||  |||  ||  |||| ||||||  |       |   |
Db       1049 TTGCACAAAATGTTACCATAGAAAACGCAATTGGTGGTTTTGGTAATGATATTATTCATG 1108

Qy        245 GCCAGGGTGCGGATAATAGTACTATTGAA 273
              |  | | |||  |||||||  |  |||||
Db       1109 GTAATGATGCTGATAATACTCTTATTGGA 1137


RESULT 15
US-08-913-159-12
; Sequence 12, Application US/08913159
; Patent No. 6300109
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Plasmid-derived type II
;    TITLE OF INVENTION:  restriction-modification systems from Lactococcus lactis
;    NUMBER OF SEQUENCES:  14
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/913,159
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 0179/95
;      FILING DATE:  17-FEB-1995
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2355 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Lactococcus lactis subsp. cremoris
;      STRAIN:  W39
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..1283
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 744
;      OTHER INFORMATION:  /product= "LlaDII restriction endonuclease"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORF"
;      OTHER INFORMATION:  /number= 1
;      OTHER INFORMATION:  /standard_name= "Gene coding for R.LlaDII"
;      OTHER INFORMATION:  /label= r-llaDII
;      OTHER INFORMATION:  /note= "The first ten amino acids in this sequence may be
;      OTHER INFORMATION:  doubtful. However, from base 773 this reading frame gives a h
;      OTHER INFORMATION:  homology with the Bsp6I endonuclease"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1392..2342
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1392
;      OTHER INFORMATION:  /product= "LlaDII methylase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "Gene coding for M.LlaDII"
;      OTHER INFORMATION:  /label= m-llaDII
;      OTHER INFORMATION:  /note= "The sequence shows 60 % identity and 76 % similarity
;      OTHER INFORMATION:  with the Bsp6I methylase."
US-08-913-159-12

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2355;
  Best Local Similarity   52.7%;  Pred. No. 1.5;
  Matches   68;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy        171 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAA 230
              || ||||    || |  |  || ||  |  ||| |   |  |   ||| |||||| || |
Db       1577 TTTTGATATTATGTTAGCAGGTTTTCCTTGCCAAGCCTTTTCTATTGCTGGTTATCGTCA 1636

Qy        231 CGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTT 290
              |||    | |  |  || |||   | ||||  |  | |||||||   ||  | |
Db       1637 AGGCTTTAACGATGAACAAGGTCGAGGTAATCTTTTTTTTGAACTTGTTCGTATTTTAGA 1696

Qy        291 CAGAAATAA 299
              | ||| ||
Db       1697 AACAAAAAA 1705


Search completed: March 15, 2004, 10:14:47
Job time : 58.8 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 07:24:24 ; Search time 340.6 Seconds
(without alignments)
4928.974 Million cell updates/sec

Title: US-09-543-407-25
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	36	7.9	506	14	US-10-029-386-20619	Sequence 20619, A
2	36	7.9	1168	14	US-10-017-161-2179	Sequence 2179, Ap
3	36	7.9	1168	15	US-10-292-798-1825	Sequence 1825, Ap
c 4	36	7.9	2428	15	US-10-108-260A-320	Sequence 320, App
c 5	34	7.5	1254	15	US-10-369-493-37688	Sequence 37688, A
6	33.8	7.4	2547	12	US-10-282-122A-36154	Sequence 36154, A
7	33.4	7.3	1355	15	US-10-045-674-593	Sequence 593, App
c 8	33.2	7.3	392	12	US-10-282-122A-25231	Sequence 25231, A
9	32.2	7.1	715	12	US-10-424-599-10132	Sequence 10132, A
c 10	32.2	7.1	849	14	US-10-032-585-6746	Sequence 6746, Ap
c 11	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
12	32.2	7.1	2266	15	US-10-108-260A-375	Sequence 375, App
c 13	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
14	32	7.0	1340	12	US-10-424-599-21469	Sequence 21469, A
c 15	32	7.0	1344	9	US-09-835-684-6	Sequence 6, Appli
c 16	32	7.0	1344	9	US-09-880-371-6	Sequence 6, Appli
c 17	32	7.0	1344	9	US-09-879-248-5	Sequence 5, Appli
c 18	32	7.0	1344	14	US-10-010-390-6	Sequence 6, Appli
c 19	32	7.0	1344	15	US-10-441-736-5	Sequence 5, Appli
20	32	7.0	14770	10	US-09-873-367C-230	Sequence 230, App
21	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
22	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
23	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c 24	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
c 25	31.8	7.0	139389	15	US-10-236-031B-61	Sequence 61, Appl
c 26	31.6	6.9	1593	12	US-10-389-647-9	Sequence 9, Appli
c 27	31.6	6.9	1593	14	US-10-127-032-32	Sequence 32, Appl
28	31.4	6.9	51259	15	US-10-374-077-209	Sequence 209, App
c 29	31.2	6.8	512	14	US-10-029-386-6908	Sequence 6908, Ap
c 30	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
31	31.2	6.8	1451	12	US-10-424-599-93896	Sequence 93896, A
32	31.2	6.8	2940917	15	US-10-027-632-174763	Sequence 174763,
33	31.2	6.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
34	31	6.8	1056	12	US-10-282-122A-12689	Sequence 12689, A
c 35	31	6.8	1513	12	US-10-425-114-28565	Sequence 28565, A
36	31	6.8	2751	15	US-10-398-221-3563	Sequence 3563, Ap
c 37	30.8	6.8	548	15	US-10-027-632-287621	Sequence 287621,
c 38	30.8	6.8	766	9	US-09-864-761-19608	Sequence 19608, A
c 39	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
c 40	30.8	6.8	1944	9	US-09-864-761-2825	Sequence 2825, Ap
41	30.8	6.8	1959	9	US-09-864-761-4012	Sequence 4012, Ap
42	30.8	6.8	3035	9	US-09-881-752A-279	Sequence 279, App
43	30.6	6.7	1567	16	US-10-389-566-344	Sequence 344, App
44	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appli
c 45	30.6	6.7	1814	15	US-10-369-493-26830	Sequence 26830, A

ALIGNMENTS

RESULT 1
US-10-029-386-20619/c
; Sequence 20619, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR (
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20619
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002070.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.79
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
US-10-029-386-20619

Query Match 7.9%; Score 36; DB 14; Length 506;
Best Local Similarity 51.9%; Pred. No. 0.15;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        149 ACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAA 208
              | |||   | | | | |||   | ||||| |    | ||   || || |    ||||
Db        452 ATGGTGATGATGATGATGCTGGTGATGATGATGGTGATAATGATGGTGATGGTGATGATG 393

Qy        209 TGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
```

```
Db        392 GTGATGATGGTGGTGATGATGATGGTGATGATGATGGTGATGGTGAGGATGGTGGTGATG 333

Qy        269 TTGAACTGACTCAGAATGGTTTCAGAAATAATGCCA 304
              ||        | |  |||||     |  || ||| ||
Db        332 ATGGTGGTGATGATGATGGTGATGGTGATGATGACA 297


RESULT 2
US-10-017-161-2179
; Sequence 2179, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2179
;   LENGTH: 1168
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1168)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(968)
US-10-017-161-2179

  Query Match             7.9%;  Score 36;  DB 14;  Length 1168;
  Best Local Similarity   51.9%;  Pred. No. 0.24;
  Matches   81;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        149 ACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAA 208
              | |||   | | | | |||   | ||||| |    | ||    || || |    ||||
Db        285 ATGGTGATGATGATGATGCTGGTGATGATGATGGTGATAATGATGGTGATGGTGATGATG 344

Qy        209 TGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |   |||  ||| ||| | | || |  ||||  ||  | |||| |||| | ||  |
Db        345 GTGATGATGGTGGTGATGATGATGGTGATGATGATGGTGATGGTGAGGATGGTGGTGATG 404

Qy        269 TTGAACTGACTCAGAATGGTTTCAGAAATAATGCCA 304
              ||        | |  |||||     |  || ||| ||
Db        405 ATGGTGGTGATGATGATGGTGATGGTGATGATGACA 440


RESULT 3
US-10-292-798-1825
; Sequence 1825, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1825
;   LENGTH: 1168
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1168)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(968)
US-10-292-798-1825

  Query Match             7.9%;  Score 36;  DB 15;  Length 1168;
  Best Local Similarity   51.9%;  Pred. No. 0.24;
  Matches   81;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        149 ACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAA 208
              | |||   | | | | |||   | ||||| |    | ||    || || |    ||||
Db        285 ATGGTGATGATGATGATGCTGGTGATGATGATGGTGATAATGATGGTGATGGTGATGATG 344

Qy        209 TGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |   |||  ||| ||| | | || |  ||||  ||  | |||| |||| | ||  |
Db        345 GTGATGATGGTGGTGATGATGATGGTGATGATGATGGTGATGGTGAGGATGGTGGTGATG 404

Qy        269 TTGAACTGACTCAGAATGGTTTCAGAAATAATGCCA 304
              ||        | |  |||||     |  || ||| ||
Db        405 ATGGTGGTGATGATGATGGTGATGGTGATGATGACA 440


RESULT 4
US-10-108-260A-320/c
; Sequence 320, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 320
;   LENGTH: 2428
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-320

  Query Match             7.9%;  Score 36;  DB 15;  Length 2428;
  Best Local Similarity   51.2%;  Pred. No. 0.36;
  Matches   84;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        136 AGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTT 195
              | | |  ||||  |  |||  |     || | || ||  |||||| |||  |   |
Db        964 ATCTTCAATCAGCCGGACTCCAATGTTAAAGCCCATTTTGTCCAGCTGTTTATTCACATC 905

Qy        196 GTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCG 255
              | | |  |   || |  | | ||      ||| ||   | |    |||| |||||||
Db        904 TTCATCATTTTCATAGTCCTTACATAGCTGGGTGACCAGGGCACCATAGGTCAGGGTGAA 845

Qy        256 GATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAA 299
              ||    ||  || |  ||  || ||| || |   |||| |  ||
Db        844 GAGCTCAGAGCTCTAAAAGAGATTCAAAAGGCAGTCAGGACCAA 801


RESULT 5
US-10-369-493-37688/c
; Sequence 37688, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
```

```
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 37688
;   LENGTH: 1254
;   TYPE: DNA
;   ORGANISM: Pseudomonas fluorescens
US-10-369-493-37688

  Query Match             7.5%;  Score 34;  DB 15;  Length 1254;
  Best Local Similarity   45.3%;  Pred. No. 1.2;
  Matches  124;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy         97 AACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCC 156
              ||||||||||      ||     ||   |||         |||     |||| ||| | |
Db        573 AACGGCGGCGCGGTCGGTGGTTTCCACCACTTGCACAACGGCAGCCGGCAGTTCGGCTGC 514

Qy        157 GCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 216
              |   | | |||||  |||         | | |   ||||  |   ||    | | |
Db        513 GGCCAGACCGCGCTGGATGCAGGCAGCGATCGCACGGTTGGAATGAATCGCCTCGGAACC 454

Qy        217 GCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              || |   | | |    |||     | |   ||||    || |    | |    ||  |
Db        453 GCCGCGCAGGATGGTCGCGTTACCGGACTTCAGGCACAGGCTGGCGGCATCGATGGTCAC 394

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
                |  ||  || |||  | || || ||     || |||| || ||||  |     ||||
Db        393 GTTGGGACGGGATTCGTAGATGATCCCGATCACGCCCAGCGGCACGCGCATCTTGCCGAC 334

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              | | || | | |     |  | | ||| | | |
Db        333 CTGAATGCCCGACGGACGGAAGCTCATATCACGG 300


RESULT 6
US-10-282-122A-36154
; Sequence 36154, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36154
;   LENGTH: 2547
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
US-10-282-122A-36154

  Query Match             7.4%;  Score 33.8;  DB 12;  Length 2547;
  Best Local Similarity   77.4%;  Pred. No. 2;
  Matches   41;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        178 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAA 230
              |||| |||| | | |||||| |||||||| |||| |||  | | | ||||||
Db        850 CAGCAGGTTGCACTTGTTGTTGCCCATGAACTGGCGCATCAATGGTTTGGTAA 902


RESULT 7
US-10-045-674-593
; Sequence 593, Application US/10045674
; Publication No. US20030232333A1
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 60/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674-593

  Query Match             7.3%;  Score 33.4;  DB 15;  Length 1355;
  Best Local Similarity   28.9%;  Pred. No. 2;
  Matches   61;  Conservative   52;  Mismatches   98;  Indels    0;  Gaps    0;

Qy         57 GGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTC 116
              :| :||:| :  :     : |  ::||:||:||:: : |   :|| ||: :: |  |:
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        117 CGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGA 176
              :||: : |       :   |:: :  : | | :| :| :|  ||    ||   :  : |
```

```
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        177 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGC 236

Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        237 CGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCT 978


RESULT 8
US-10-282-122A-25231/c
; Sequence 25231, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25231
;   LENGTH: 392
;   TYPE: DNA
;   ORGANISM: Legionella pneumophila
US-10-282-122A-25231

  Query Match             7.3%;  Score 33.2;  DB 12;  Length 392;
  Best Local Similarity   50.0%;  Pred. No. 1.2;
  Matches   83;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy        194 TTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTG 253

Db        184 TTTTTATCAATTAAATTGCGCTGAACGTAGTCGGGAAATTCCCTGAAATTTTCTATGTCA 125

Qy        254 CGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACC 313

Db        124 GGATTAATTGCATTCTTGAAAGTACCCAGTTCGTATTCATAATCATTGCGGGTGTCTAAT 65

Qy        314 AGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGG 359

Db         64 AAAATAACATTAAGATCTTGAAATGAACTGGATGCCATTCCCTCAG 19


RESULT 9
US-10-424-599-10132
; Sequence 10132, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 10132
;   LENGTH: 715
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(715)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_10915C.1
US-10-424-599-10132

  Query Match             7.1%;  Score 32.2;  DB 12;  Length 715;
  Best Local Similarity   46.6%;  Pred. No. 3.6;
  Matches  103;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy        178 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCC 237

Db        185 CTGCTTGCGGCCGGTGTTGCCAACGCCCAATTCGGTGGTCAAGGTGGTGGTCAAGGCCAG 244

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297

Db        245 AACAACGGCCAGAATGGACAAAATGGCGGCAACAATGGCCAGAACAACGGTCAAAACAAC 304

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357

Db        305 GGTCAAAACGGAGGCAACAACGGCGGTAACAACGGTGGCAACAATGGTGGTAACAACGGT 364

Qy        358 GGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398

Db        365 GGTAATGGTGGGTGGTTTGGCCCTCAATGCCGATGTCGTTC 405


RESULT 10
US-10-032-585-6746/c
; Sequence 6746, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6746
;   LENGTH: 849
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-10-032-585-6746
```

```
  Query Match             7.1%;  Score 32.2;  DB 14;  Length 849;
  Best Local Similarity   50.3%;  Pred. No. 4;
  Matches   79;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTG 183
              ||  ||'|  ||  ||  ||   ||| | | |  |||   |  || |||  | |   ||
Db        794 GATGCATCACTGGCCAAATAAAGGTATGCTCCGACTAGTTCATCGGTTTCAGCTTCTCTT 735

Qy        184 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTA 243
                  | | |||    ||||  |  | || | |   |||| ||||  ||  ||||
Db        734 CCCATCGGAGTTAAACCCCACCATTTTGCCCTTTGTGGTTTTGGTATTGGTTCCGATAAT 675

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTC 280
                |   |  ||   |     |   ||  ||| ||||||
Db        674 AGGGTGTTTGTATAACCAGGGGATACAGAATTGACTC 638


RESULT 11
US-10-369-493-33133/c
; Sequence 33133, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33133
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Xylella fastidiosa
US-10-369-493-33133

  Query Match             7.1%;  Score 32.2;  DB 15;  Length 1083;
  Best Local Similarity   59.1%;  Pred. No. 4.5;
  Matches   55;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTG 421
              ||  | ||| |||||   |  | |||||     ||  || | ||| | ||| ||||
Db        488 ATCTCACCGAGCTGGGATAAAAAACCGCGCACAATCACACCATAACGCTGCTTCAGCCAC 429

Qy        422 GTTTTGGCAACAACGCCACGGCTAACCAGTATT 454
                |||| ||||| || |||    | ||  | |||
Db        428 TTTTTAGCAACCACCCCAGCAGTCACACGCATT 396


RESULT 12
US-10-108-260A-375
; Sequence 375, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 375
;   LENGTH: 2266
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-375

  Query Match             7.1%;  Score 32.2;  DB 15;  Length 2266;
  Best Local Similarity   51.0%;  Pred. No. 6.7;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        161 ACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAG 220
              | | | | || | ||    | |  |  |   | | | || || |||  ||  ||    |
Db       1838 AAGGTTCCCTGTCTGGGTCGATAAATGGCATGGATATCACACAGGAAGGGGTGCACTTTG 1897

Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTC 280
                  |||       | | || | | | | | |  |||| ||| |     |||| ||||||
Db       1898 TCACAGGTGGAAATGACCATCTGGTCAAAGTTTGGGATTATAATGAGGGTGAAGTGACTC 1957

Qy        281 AGAATGGTTTCAGAAATAATGCCACCATC 309
                |   |||  |  || | | || || ||||
Db       1958 ACGTTGGGGTGGGACACAGTGGCAACATC 1986


RESULT 13
US-10-282-122A-27869/c
; Sequence 27869, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27869
;   LENGTH: 4830
;   TYPE: DNA
;   ORGANISM: Mycobacterium leprae
US-10-282-122A-27869

  Query Match             7.1%;  Score 32.2;  DB 12;  Length 4830;
  Best Local Similarity   51.0%;  Pred. No. 10;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        224 ATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGA 283
              ||| ||     | |||  | | | |   |||     |    | |  ||| | |    ||
```

```
Db       4229 ATGATAGTAAAGACGACTTCGCCGACATTGCCTCCGACCAGAATCGTGACCGCATCACGA 4170

Qy        284 ATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTG 343

Db       4169 ACGCCTGCCCACATGCTGCGGCCTTCGACCAGGGCATCGAGCAGTACACCCAGATCGTCG 4110

Qy        344 TCGGCCAATACGGCGGTAATAACGCCGCG 372

Db       4109 TCGGTCAAGACGATGTCAGCGGCTCCTCG 4081
```

RESULT 14
US-10-424-599-21469
; Sequence 21469, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 21469
; LENGTH: 1340
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_119391C.1
US-10-424-599-21469

Query Match 7.0%; Score 32; DB 12; Length 1340;
Best Local Similarity 54.2%; Pred. No. 6;
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy        197 TTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGG 256

Db        814 TTTCTCATCCAAGTGCACATGTCCGTGCTCTGAGCAGTTCGGTTATTCGTGACATTTTGC 873

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316

Db        874 ATACTAGTTCAATTGGATCGAGTCCTAAGCCACTGCAAAAGAATGGCATCCATGATCAGT 933
```

RESULT 15
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.0%; Score 32; DB 9; Length 1344;
Best Local Similarity 54.2%; Pred. No. 6;
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111

Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171

Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90
```

Search completed: March 15, 2004, 14:16:24
Job time : 346.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:52:04 ; Search time 2500.2 Seconds
(without alignments)
6417.435 Million cell updates/sec

Title: US-09-543-407-25
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-25	Sequence 25, Appl
2	404.8	88.8	456	6	US-08-233-642A-56	Sequence 56, Appl
3	404.8	88.8	456	23	US-09-543-407-1	Sequence 1, Appli
4	385.6	84.6	456	23	US-09-543-407-19	Sequence 19, Appl
5	355.2	77.9	456	23	US-09-543-407-17	Sequence 17, Appl
6	353.6	77.5	456	23	US-09-543-407-11	Sequence 11, Appl
7	353.6	77.5	456	23	US-09-543-407-13	Sequence 13, Appl
8	353.6	77.5	456	23	US-09-543-407-23	Sequence 23, Appl
9	352	77.2	456	23	US-09-543-407-15	Sequence 15, Appl
10	350.4	76.8	456	23	US-09-543-407-21	Sequence 21, Appl
11	350.4	76.8	456	23	US-09-543-407-27	Sequence 27, Appl
12	347.2	76.1	456	23	US-09-543-407-29	Sequence 29, Appl
13	280.6	61.5	361	6	US-08-233-642A-54	Sequence 54, Appl
14	227.2	49.8	477	17	US-09-252-691-172	Sequence 172, App
15	227.2	49.8	477	17	US-09-252-691C-172	Sequence 172, App
16	227.2	49.8	477	51	US-10-417-886-172	Sequence 172, App
17	220.8	48.4	456	23	US-09-543-407-3	Sequence 3, Appli
18	220.8	48.4	675	13	US-08-978-878-3	Sequence 3, Appli
19	220.8	48.4	675	31	US-09-741-873B-3	Sequence 3, Appli
20	218.4	47.9	453	92	US-60-352-946-1	Sequence 1, Appli
21	218.4	47.9	453	101	US-60-444-371-1	Sequence 1, Appli
22	176.8	38.8	396	13	US-08-978-878-1	Sequence 1, Appli
23	152	33.3	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
c 28	50.4	11.1	78	23	US-09-543-407-52	Sequence 52, Appl
c 29	49.6	10.9	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
30	48.2	10.6	78	23	US-09-543-407-53	Sequence 53, Appl
31	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
32	43.4	9.5	78	23	US-09-543-407-49	Sequence 49, Appl
33	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
c 34	40.6	8.9	78	23	US-09-543-407-48	Sequence 48, Appl
35	40	8.8	100	45	US-10-146-492B-79	Sequence 79, Appl
36	37.4	8.2	176594	49	US-10-322-281-495	Sequence 495, App
37	37	8.1	26029	64	US-60-082-302-707	Sequence 707, App
38	36.6	8.0	448	28	US-09-669-817A-17239	Sequence 17239, A
39	36.6	8.0	456	23	US-09-543-407-4	Sequence 4, Appli
c 40	36.6	8.0	515	28	US-09-669-817A-20515	Sequence 20515, A
c 41	36.6	8.0	546	27	US-09-654-617-416236	Sequence 416236,
c 42	36.6	8.0	546	28	US-09-669-817A-25010	Sequence 25010, A
c 43	36.6	8.0	546	29	US-09-684-016-416236	Sequence 416236,
c 44	36.6	8.0	1253	26	US-09-620-392-64068	Sequence 64068, A
c 45	36.6	8.0	1253	30	US-09-702-134-3537	Sequence 3537, Ap

ALIGNMENTS

```
RESULT 1
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 5.9e-133;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;     APPLICANT:  Kay, William W.
;     APPLICANT:  Collinson, S. Karen
;     APPLICANT:  Clouthier, Sharon C.
;     APPLICANT:  Doran, James L.
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;     TITLE OF INVENTION:  BASED VACCINES
;     NUMBER OF SEQUENCES:  58
;     NUMBER OF SEQUENCES:  -
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Seed and Berry
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  U.S.A.
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/233,642A
;       FILING DATE:  26-APR-1994
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  King, Joshua
```

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 9.2e-117;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||   |    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 9.2e-117;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||   |    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             84.6%;  Score 385.6;  DB 23;  Length 456;
  Best Local Similarity   90.4%;  Pred. No. 1.1e-110;
  Matches  412;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 4.5e-101;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 77.5%; Score 353.6; DB 23; Length 456;
Best Local Similarity 86.0%; Pred. No. 1.4e-100;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 1.4e-100;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||      |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||        |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 1.4e-100;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||      |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||      ||  ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                || |    || |  | ||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 4.6e-100;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | | |  |  |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||    |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 1.5e-99;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |    ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||    |          ||  | ||    ||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  | |  | || |   || | |   | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 1.5e-99;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 76.1%; Score 347.2; DB 23; Length 456;
Best Local Similarity 85.1%; Pred. No. 1.5e-98;
Matches 388; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 61.5%; Score 280.6; DB 6; Length 361;
Best Local Similarity 89.9%; Pred. No. 1.6e-77;
Matches 301; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123

Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTG 183

Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTA 243
              | | |||||     |          ||  | ||   ||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 14
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-172

Query Match 49.8%; Score 227.2; DB 17; Length 477;
Best Local Similarity 68.6%; Pred. No. 1.4e-60;
Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| |||||||||    | |||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||     |  |  | || |||  | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              || ||||||||  ||||||||| ||||||||   || |||| | ||||||  |   |||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                ||   |||| | |  |          ||  | ||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||     ||| |||| | ||| |  ||    || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||||  |||  | ||||| || || |  || |  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  |||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477
```

RESULT 15
US-09-252-691C-172

; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691C
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-172

Query Match 49.8%; Score 227.2; DB 17; Length 477;
Best Local Similarity 68.6%; Pred. No. 1.4e-60;
Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| |||||||||    | |||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||     |  |  | || |||  | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              || ||||||||  ||||||||| ||||||||   || |||| | ||||||  |   |||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                ||   |||| | |  |          ||  | ||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||     ||| |||| | ||| |  ||    || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||||  |||  | ||||| || || |  || |  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  |||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477
```

Search completed: March 15, 2004, 13:42:49
Job time : 2501.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 04:16:29 ; Search time 53.4 Seconds
(without alignments)
2945.701 Million cell updates/sec

Title: US-09-543-407-25
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 628836 seqs, 172478556 residues

Total number of hits satisfying chosen parameters: 1257672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	220.8	48.4	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	176.8	38.8	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33.4	7.3	1355	6	US-10-045-674A-593	Sequence 593, App
c	4	31.8	7.0	290892	6	US-10-767-471-10809	Sequence 10809, A
c	5	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
	6	30.8	6.8	3035	6	US-10-786-892-279	Sequence 279, App
c	7	30.8	6.8	803755	6	US-10-767-471-10700	Sequence 10700, A
	8	29.8	6.5	1588	6	US-10-767-701-14522	Sequence 14522, A
c	9	29.8	6.5	79977	6	US-10-765-790-58	Sequence 58, Appl
	10	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
	11	29.4	6.4	2205	6	US-10-767-701-14689	Sequence 14689, A
c	12	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	13	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
	14	28.8	6.3	1963	6	US-10-100-683-5215	Sequence 5215, Ap
	15	28.8	6.3	2133	6	US-10-100-683-5214	Sequence 5214, Ap
	16	28.8	6.3	2166	6	US-10-100-683-2906	Sequence 2906, Ap
	17	28.8	6.3	2166	6	US-10-100-683-2907	Sequence 2907, Ap
c	18	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	19	28.6	6.3	596	6	US-10-767-701-5378	Sequence 5378, Ap
	20	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	21	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	22	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
c	23	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	24	28.2	6.2	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	25	28.2	6.2	3416	6	US-10-100-683-3397	Sequence 3397, Ap
	26	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
	27	28	6.1	752	6	US-10-767-795-3801	Sequence 3801, Ap
	28	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
	29	27.8	6.1	861	6	US-10-767-701-4852	Sequence 4852, Ap
c	30	27.8	6.1	1017	6	US-10-771-241-213	Sequence 213, App
	31	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
c	32	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	33	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
	34	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
	35	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
c	36	27.6	6.1	1079	6	US-10-767-701-14615	Sequence 14615, A
	37	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	38	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	39	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	40	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	41	27.4	6.0	609	6	US-10-767-701-23877	Sequence 23877, A
c	42	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
c	43	27.4	6.0	1686	6	US-10-767-701-6632	Sequence 6632, Ap
	44	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
	45	27.2	6.0	616	6	US-10-767-701-26114	Sequence 26114, A

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 48.4%; Score 220.8; DB 5; Length 675;
Best Local Similarity 67.8%; Pred. No. 2.3e-65;
Matches 309; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| |||  |   ||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |  |       ||  | ||    ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 38.8%; Score 176.8; DB 5; Length 396;
Best Local Similarity 65.4%; Pred. No. 1.6e-50;
Matches 259; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| |||  |   ||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |  |       ||  | ||    ||    ||||| || || |||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
```

```
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
US-10-045-674A-593

Query Match 7.3%; Score 33.4; DB 6; Length 1355;
Best Local Similarity 28.9%; Pred. No. 0.16;
Matches 61; Conservative 52; Mismatches 98; Indels 0; Gaps 0;

```
Qy         57 GGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTC 116
              :| :||:| :  :     : |  ::||:||:||:: : |   :|| ||: :: |  |:
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        117 CGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGA 176
              :||: : |       :   |:: :  : | | :| :| :|  ||    ||   :  : |
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        177 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGC 236
              : |:  :| :|    :|    : |: |:||:|  ||:|    : :| |:| :|| ||:
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        237 CGATGTAGGCCAGGGTGCGGATAATAGTACT 267
              :|   ::   | : | ||: |: || ||
Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCT 978
```

RESULT 4
US-10-767-471-10809/c

```
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
;   LENGTH: 290892
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(290892)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

  Query Match             7.0%;  Score 31.8;  DB 6;  Length 290892;
  Best Local Similarity   46.0%;  Pred. No. 6.5;
  Matches  108;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        149 ACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAA 208
Db     234486 ACGATGATGGTTATGGTTACAGTTATGATGTTGATGTTGGTGATGATGGTGGTGGTGATG 234427

Qy        209 TGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
Db     234426 ATGATGGTGATGATAATGATTATGATGATGGTGTTGGTAATGGTGATGATGGTGATGATA 234367

Qy        269 TTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAA 328
Db     234366 ATGACAATGATAAGGATGGTAATGAAAATGATAATGCTAATGCTAATTGTGATGATAATG 234307

Qy        329 ACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCA 383
Db     234306 GAAACCATGGCAATGATGATGGTAACTATCGTGATAATGATGATGATAATGA 234252


RESULT 5
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
;  APPLICANT: Bayer Pharmaceuticals Corporation
;  APPLICANT:  Eveleigh, Deepa
;  APPLICANT:  Bigwood, Douglas
;  TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
;  FILE REFERENCE: 5152
;  CURRENT APPLICATION NUMBER: US/10/788,792
;  CURRENT FILING DATE:  2004-02-27
;  PRIOR APPLICATION NUMBER: US 60/450,655
;  PRIOR FILING DATE: 2003-02-28
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
;   LENGTH: 1575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-788-792-27

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 1575;
  Best Local Similarity   55.7%;  Pred. No. 1.3;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583


RESULT 6
US-10-786-892-279
; Sequence 279, Application US/10786892
; GENERAL INFORMATION:
;  APPLICANT: Kleanthous, Harold
;  APPLICANT:  Al-Garawi, Amal
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Tomb, Jean-Francois
;  APPLICANT:  Oomen, Raymond P.
;  TITLE OF INVENTION: Identification of Polynucleotides
;  TITLE OF INVENTION:  Encoding Novel Helicobacter Polypeptides in the Helicobacter
;  TITLE OF INVENTION:  Genome
;  FILE REFERENCE: 06132/041003
;  CURRENT APPLICATION NUMBER: US/10/786,892
;  CURRENT FILING DATE:  2004-02-25
;  PRIOR APPLICATION NUMBER: US 09/881,752
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: US 08/833,457
;  PRIOR FILING DATE: 1997-04-01
;  NUMBER OF SEQ ID NOS: 370
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 279
;   LENGTH: 3035
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (729)...(2981)
US-10-786-892-279

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 3035;
  Best Local Similarity   46.3%;  Pred. No. 1.8;
  Matches  101;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy        166 GCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTAT 225
Db        180 GAGCAGTTAGAAAAGCTTAATGATCATTCTTTCACGCCTAAAGAATGGGACACTCTTTAT 239

Qy        226 GGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAAT 285
Db        240 TCTCAATTCATCGCTAATAAAAACGATGACTATAAGGCTAAAACGAAAAAGATCCAAGAA 299

Qy        286 GGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTC 345
Db        300 GATCCGATTTTTAATCTCACGCTAGAGAACGGGAAAACCAAAAACATTAAAATCATTGAT 359

Qy        346 GGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCA 383
Db        360 AAGAAAAATATCCATAGAAACGCCTTGCAAGTGATCCA 397


RESULT 7
US-10-767-471-10700/c
; Sequence 10700, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10700
;   LENGTH: 803755
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(803755)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
```

US-10-767-471-10700

Query Match 6.8%; Score 30.8; DB 6; Length 803755;
Best Local Similarity 57.1%; Pred. No. 22;
Matches 56; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy        172 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAAC 231
              ||  ||| | |  | ||   |  |||  || || ||| |||||| | ||| | |||
Db     738590 TAATGTCATCAGTCAATCCTGTTATTTATTCAGGACATGACACATGAATGTTGTTGTACT 738531

Qy        232 GGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTAT 269
              ||   | | |||||  |  || | ||| ||   | ||
Db     738530 GGTATCATTCTAGGCATGCATGAGAATACTAAGGCAAT 738493
```

RESULT 8
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
; LENGTH: 1588
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

Query Match 6.5%; Score 29.8; DB 6; Length 1588;
Best Local Similarity 44.1%; Pred. No. 2.9;
Matches 124; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

```
Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
              | |  |||| |        | ||| | | |       | ||||| |||  |  |  |  |
Db        455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
              ||||| ||| || ||   |||    |     |  |   |||           ||||  |  |
Db        515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy        160 AACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 219
                | ||   ||  |||| |     | | |  |||  | |        |  | ||   || |
Db        575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              |||  |   | |  |||        |   ||||| || | | ||  |  || |   ||
Db        635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAA 320
                | | ||  |   |  | ||||  || | ||  | ||
Db        695 GCTGGTAGCTTTGGCGGTGGTTCCACTGGCGCCGAGAAGGA 735
```

RESULT 9
US-10-765-790-58/c
; Sequence 58, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
; LENGTH: 79977
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-58

Query Match 6.5%; Score 29.8; DB 6; Length 79977;
Best Local Similarity 45.5%; Pred. No. 17;
Matches 106; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
              | |  |||  || |  || ||| |  | |    |  |||  |  || ||| || || |
Db      14910 GATGGTGGTGATGGTGAGGATGGTGATGGTGATGATGGGGATGATGGTGGTGATGATGAT 14851

Qy         100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
               | || ||  || ||   |      ||                |  | | ||    | |  | |
Db      14850 GATGGTGGTGATGGTGATGATGGTGATGGTGGTGAGGATAGTGATGGTGGTGATGGTGAT 14791

Qy         160 AACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 219
               || |  |  | ||| |     | |   | || || |    |||    ||    ||
Db      14790 AATGGTTGTGATGATGGTGGTGATGGTGATCCTGATGGTGGTGATGGTGAGGATGGTGAT 14731

Qy         220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGA 272
               ||| |||||  ||  |  ||||  |  | || |  |||  | ||  |  |||
Db      14730 GGTGATGGTGGCGATGGTGATGGTGATGATGGGGATGATGGTGGTGATGATGA 14678
```

RESULT 10
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

Query Match 6.4%; Score 29.4; DB 6; Length 300;
Best Local Similarity 54.1%; Pred. No. 1.9;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy          14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
               |||  |  |  | || |||      |     ||||  | |    | |  || || ||| |
Db         123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy          74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
               |||||||||||||| ||     |||||||||  |  ||  |  || ||
Db         183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233
```

RESULT 11
US-10-767-701-14689
; Sequence 14689, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.

```
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14689
;   LENGTH: 2205
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS833_1
US-10-767-701-14689

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 2205;
  Best Local Similarity   50.3%;  Pred. No. 4.6;
  Matches   72;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        166 GCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTAT 225
              || |||| |||  |||||||| |    | | |  | | | |  |    ||||   || |
Db        921 GCTCTTTTTGACAAGCTGGTTTCAACAGCTATGGCTGAAGGATGGCTTGATGCTTCTTTT 980

Qy        226 GGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAAT 285
              |||||    || |      | | | |    | | ||||  | |  || |  ||   || ||
Db        981 GGTAAATCTGCTGCCCCTGACGAAGAGATGCAAAATACAAGTGCTGTAAAGATGAAGCAT 1040

Qy        286 GGTTTCAGAAATAATGCCACCAT 308
                |    | |   ||||  |||
Db       1041 TTTAAAGAAGAATCTGCCCACAT 1063


RESULT 12
US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-117

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 249;
  Best Local Similarity   52.5%;  Pred. No. 2;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   |||||||        || || ||| |    | |  | |   || | | |||
Db        180 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 121

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    |  |  ||  || || || ||| |  | |||    |||| | | |    || |
Db        120 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 61

Qy        415 CA 416
              |
Db         60 GA 59


RESULT 13
US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 757
;   TYPE: DNA
;   ORGANISM: E. Coli
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(757)
;   OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 757;
  Best Local Similarity   52.5%;  Pred. No. 3.3;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   |||||||        || || ||| |    | |  | |   || | | |||
Db        264 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 323

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    |  |  ||  || || || ||| |  | |||    |||| | | |    || |
Db        324 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 383

Qy        415 CA 416
              |
Db        384 GA 385


RESULT 14
US-10-100-683-5215
; Sequence 5215, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
```

```
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5215
;   LENGTH: 1963
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-5215

  Query Match             6.3%;  Score 28.8;  DB 6;  Length 1963;
  Best Local Similarity   54.8%;  Pred. No. 7;
  Matches   57;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        214 CATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
              |||| ||| || || |  || |  |||| |  |    ||||| || || |        |
Db       1316 CATGAAGGCTACGGCATTGGGGATGATGAATACTCCTGTGCGTATGATGGCTGCCGGCAG 1375

Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              ||||  |  | |    |  ||||| |||  |   |||  |||  ||
Db       1376 CTGATTTGGTACAATGCCAGAAGTAAGCCTCACATACACCCATG 1419


RESULT 15
US-10-100-683-5214
; Sequence 5214, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5214
;   LENGTH: 2133
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-5214

  Query Match             6.3%;  Score 28.8;  DB 6;  Length 2133;
  Best Local Similarity   54.8%;  Pred. No. 7.3;
  Matches   57;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        214 CATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
              |||| ||| || || |  || |  |||| |  |    ||||| || || |        |
Db       1446 CATGAAGGCTACGGCATTGGGGATGATGAATACTCCTGTGCGTATGATGGCTGCCGGCAG 1505

Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              ||||  |  | |    |  ||||| |||  |   |||  |||  ||
Db       1506 CTGATTTGGTACAATGCCAGAAGTAAGCCTCACATACACCCATG 1549
```

Search completed: March 15, 2004, 13:47:26
Job time : 55.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 02:55:53 ; Search time 1963.4 Seconds
(without alignments)
6935.497 Million cell updates/sec

Title: US-09-543-407-25
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	221.4	48.6	680	12	BJ618688	BJ618688 BJ618688
c	2	40	8.8	1201	13	BX379637	BX379637 BX379637
	3	38	8.3	636	12	BM935657	BM935657 UI-M-BH3-
	4	38	8.3	651	14	CD771910	CD771910 AGENCOURT
	5	37.4	8.2	366	9	AA672532	AA672532 vo59f11.r
	6	37.4	8.2	404	9	AA672619	AA672619 vo59h11.r
c	7	37.4	8.2	484	9	AV430994	AV430994 AV430994
	8	37.4	8.2	512	9	AA763679	AA763679 vp09h03.r
	9	37.4	8.2	713	12	BG914781	BG914781 602813566
	10	37.4	8.2	3362	11	AK040558	AK040558 Mus muscu
c	11	36.6	8.0	711	9	AU101070	AU101070 AU101070
	12	35.6	7.8	500	28	AZ180696	AZ180696 SP_0181_A
c	13	35.4	7.8	664	13	CA141136	CA141136 SCJFRT205
	14	35	7.7	556	28	AQ660801	AQ660801 Sheared D
	15	35	7.7	679	28	AQ952353	AQ952353 Sheared D
c	16	35	7.7	764	14	CA752110	CA752110 UI-M-FD0-
c	17	34.6	7.6	358	10	AW487009	AW487009 78612 MAR
c	18	34.6	7.6	420	14	CD286448	CD286448 12_B14.ab
c	19	34.6	7.6	480	12	BM364426	BM364426 BS3200010
c	20	34.6	7.6	513	12	BM363458	BM363458 BS320057A
c	21	34.6	7.6	559	13	BU765156	BU765156 sas23f11.
c	22	34.6	7.6	564	10	AW632655	AW632655 92714 MAR
	23	34.4	7.5	201	29	CE815903	CE815903 tigr-gss-
	24	34.4	7.5	715	13	BU444683	BU444683 603765715
	25	34.2	7.5	296	9	AV058630	AV058630 AV058630
c	26	34.2	7.5	479	28	BH211305	BH211305 Sm1-50G22
	27	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
	28	34.2	7.5	3470	11	AK012196	AK012196 Mus muscu
c	29	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	30	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
	31	33.8	7.4	468	12	BG553005	BG553005 dab82e09.
c	32	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	33	33.8	7.4	1893	29	AY414592	AY414592 Pan trogl
	34	33.6	7.4	508	28	AQ846320	AQ846320 LMAJFV1_1
	35	33.6	7.4	519	12	BJ275419	BJ275419 BJ275419
c	36	33.6	7.4	519	28	AZ085047	AZ085047 RPCI-23-3
c	37	33.6	7.4	569	12	BJ270411	BJ270411 BJ270411
	38	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	39	33.6	7.4	606	12	BJ240656	BJ240656 BJ240656
	40	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
	41	33.6	7.4	769	14	CB501642	CB501642 ssalga504
c	42	33.4	7.3	505	6	AL809250	Al809250 Triticum
	43	33.4	7.3	583	13	BU041698	BU041698 PP_LEa001
	44	33.4	7.3	688	13	BU026576	BU026576 QHG17F15.
	45	33.4	7.3	707	13	BU023379	BU023379 QHF10N10.

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 48.6%; Score 221.4; DB 12; Length 680;
Best Local Similarity 67.8%; Pred. No. 2.3e-51;
Matches 309; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||  | ||   | ||| |||||| |||||||||  || ||||  ||| |||  |    ||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |  |       ||  | ||    ||    |||||  || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
BX379637/c
LOCUS BX379637 1201 bp mRNA linear EST 08-MAY-2003
DEFINITION BX379637 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI036YK14 5-PRIME, mRNA sequence.
ACCESSION BX379637
VERSION BX379637.1 GI:30450782
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DI036BF07QP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI036YK14"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 8.8%; Score 40; DB 13; Length 1201;
Best Local Similarity 26.7%; Pred. No. 3.7;
Matches 63; Conservative 74; Mismatches 99; Indels 0; Gaps 0;

```
Qy          8 TTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCG 67
              :|| |:   ::   : | |  :  || ||:  :|||:| : :: :  :|  :::|:::::
Db       1064 YTTAAMTTTYSWTTBTACTTCSTCCATTCSCGSTTTYTCSGHBCSTCSTTTYBSGSKDBS 1005

Qy         68 TTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACT 127
              :|:: : | |::|::: : :      :|:  |: :|: : :  | ::|:::    |
Db       1004 KTYSGSKACGBSGSSSKTSABTCSTBASADTKCRTSGSASTSBTTTSDTBBSTTTTGASGA 945

Qy        128 CAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTGGTTA 187
              | |::  :  |: :| ::| |:|: || |  :|   :  |: |:| | | ||  : :| |
Db        944 CGAYSCSSTTCRCKTYYVAATMCSTTTTCTTHACGASACCSTTYTCTTAACAATYCSTCA 885

Qy        188 CCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTA 243
                ||:| :  || |:  |: | | :|  : |   : ||  : | :|: : :: | :||
Db        884 TCCBTTYCTTTTCSTTTBCAYTCBCSYSTMACVATTTCVTTSMCSTYTYYCACBTA 829
```

RESULT 3
BM935657
LOCUS BM935657 636 bp mRNA linear EST 13-MAR-2002
DEFINITION UI-M-BH3-arw-e-09-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone UI-M-BH3-arw-e-09-0-UI 5', mRNA sequence.
ACCESSION BM935657
VERSION BM935657.1 GI:19394809
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 636)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA sequence: 245-280, >AT_rich#Low_complexity 459-514, >(CAT)n#Simple_repeat (matched compliment)
Seq primer: M13 REVERSE.
FEATURES Location/Qualifiers
source 1. .636

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-arw-e-09-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_M_S4"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)"

ORIGIN

Query Match 8.3%; Score 38; DB 12; Length 636;
Best Local Similarity 53.0%; Pred. No. 10;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        175 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGC 234
              |||||| | |||  | |  |||  |   |||| || |  |||   |||  ||||   ||
Db        419 GATCAGTTTGTTTTCTGACTTGAAAGATATGAGATTGTTCATCANGGTGGTGGTGGTGGT 478

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              |  ||||  |   |   |||  |||||  |  |  | ||  ||   | |||  | |   |
Db        479 GGTGATGATGATGATCATGCTCATAATGATGGTGATTAAAAGAAAGAAAATTTTATATCA 538

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTA 325
              |||  | |     | | |||  | ||  |||
Db        539 AATGTTACACTTCTGGCCCACAGAAAGACTA 569
```

RESULT 4
CD771910
LOCUS CD771910 651 bp mRNA linear EST 02-JUL-2003
DEFINITION AGENCOURT_14736596 NIH_MGC_190 Mus musculus cDNA clone IMAGE:30504208 5', mRNA sequence.
ACCESSION CD771910
VERSION CD771910.1 GI:32430412
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 651)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Yoshihiko Yamada, Takashi Nakamura, NIDCR
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM231 row: e column: 17
High quality sequence stop: 378.
FEATURES Location/Qualifiers
source 1. .651
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30504208"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_190"
/note="Organ: Pooled - Molar; Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); Non-normalized full-length enriched library 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.71 kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA) Corp."
ORIGIN

Query Match 8.3%; Score 38; DB 14; Length 651;
Best Local Similarity 53.0%; Pred. No. 10;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        175 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGC 234
              |||||| | |||  | |  |||  |   |||| || |  |||   |||  ||||   ||
Db        265 GATCAGTTTGTTTTCTGACTTGAAAGATATGAGATTGTTCATCANGGTGGTGGTGGTGGT 324

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              |  ||||  |   |   |||  |||||  |  |  | ||  ||   | |||  | |   |
Db        325 GGTGATGATGATGATCATGCTCATAATGATGGTGATTAAAAGAAAGAAAATTTTATATCA 384

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTA 325
              |||  | |     | | |||  | ||  |||
Db        385 AATGTTACACTTCTGGCCCACAGAAAGACTA 415
```

RESULT 5
AA672532
LOCUS AA672532 366 bp mRNA linear EST 25-NOV-1997
DEFINITION vo59f11.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1054221 5', mRNA sequence.
ACCESSION AA672532
VERSION AA672532.1 GI:2644749
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 366)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:585797
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 360.

FEATURES Location/Qualifiers
source 1. .366
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1054221"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_mammary_gland_NbMMG"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 366;
Best Local Similarity 53.0%; Pred. No. 12;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        175 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGC 234
              ||||||  | |||  | |  |||  |   |||| || |  |||   |||  ||||   ||
Db         31 GATCAGTTTGTTTTCTGACTTGAAAGATATGAGATTGTTCATCAGGGTGGTGGTGGTGGT 90

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              |  ||||  |   |   |||  |||||  |  |  | ||  ||   | |||  | |   |
Db         91 GGTGATGATGATGATCATGCTCATAATGATGGTGATTAAAAGAAAGAAAATTTTATATCA 150

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTA 325
              |||  | |     | | |||  | ||  |||
Db        151 AATGTTACACTTCTGGCCCACAGAAAGACTA 181
```

RESULT 6
AA672619
LOCUS AA672619 404 bp mRNA linear EST 25-NOV-1997
DEFINITION vo59h11.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1054245 5', mRNA sequence.
ACCESSION AA672619
VERSION AA672619.1 GI:2644836
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 404)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:585821
Seq primer: -28m13 rev2 ET from Amersham.

FEATURES Location/Qualifiers
source 1. .404
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1054245"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_mammary_gland_NbMMG"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 404;
Best Local Similarity 53.0%; Pred. No. 12;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        175 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGC 234
              ||||||  | |||  | |  |||  |   |||| || |  |||   |||  ||||   ||
Db         31 GATCAGTTTGTTTTCTGACTTGAAAGATATGAGATTGTTCATCAGGGTGGTGGTGGTGGT 90

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              |  ||||  |   |   |||  |||||  |  |  | ||  ||   | |||  | |   |
Db         91 GGTGATGATGATGATCATGCTCATAATGATGGTGATTAAAAGAAAGAAAATTTTATATCA 150

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTA 325
              |||  | |     | | |||  | ||  |||
Db        151 AATGTTACACTTCTGGCCCACAGAAAGACTA 181
```

RESULT 7
AV430994/c
LOCUS AV430994 484 bp mRNA linear EST 23-AUG-2000
DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone PL027e10_r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
REFERENCE 1 (bases 1 to 484)
AUTHORS Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and Tabata,S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
PUBMED 10907854
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

```
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Porphyra yezoensis"
                     /mol_type="mRNA"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PL027e10_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             8.2%;  Score 37.4;  DB 9;  Length 484;
  Best Local Similarity   57.1%;  Pred. No. 13;
  Matches   68;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |   | |  | | ||  ||| |||||   ||| ||||| ||||||  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
               | ||  ||   ||  |||| ||| ||||||     |  ||    |||  |  || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323


RESULT 8
AA763679
LOCUS       AA763679                512 bp    mRNA    linear   EST 27-JAN-1998
DEFINITION  vp09h03.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1068149 5', mRNA sequence.
ACCESSION   AA763679
VERSION     AA763679.1  GI:2813761
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:590509
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1068149"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
ORIGIN

  Query Match             8.2%;  Score 37.4;  DB 9;  Length 512;
  Best Local Similarity   53.0%;  Pred. No. 14;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        175 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGC 234
              |||||| | |||  | |  |||  |   |||| || |  |||   |||  ||||   ||
Db         31 GATCAGTTTGTTTTCTGACTTGAAAGATATGAGATTGTTCATCAGGGTGGTGGTGGTGGT 90

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              | ||||  |   |   |||  |||||  |  |  | ||  ||   | |||  | |   |
Db         91 GGTGATGATGATGATCATGCTCATAATGATGGTGATTAAAAGAAAGAAAATTTTATATCA 150

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTA 325
              |||  | |    | | |||  | ||  |||
Db        151 AATGTTACACTTCTGGCCCACAGAAAGACTA 181


RESULT 9
BG914781
LOCUS       BG914781                713 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  602813566F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4935880 5',
            mRNA sequence.
ACCESSION   BG914781
VERSION     BG914781.1  GI:14295244
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 713)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10867  row: c  column: 17
            High quality sequence stop: 609..
FEATURES             Location/Qualifiers
     source          1. .713
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4935880"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam4"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and  Differentiation 7, 3-11 (1996)."
ORIGIN
```

```
  Query Match             8.2%;  Score 37.4;  DB 12;  Length 713;
  Best Local Similarity   53.0%;  Pred. No. 16;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        175 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGC 234
              ||||||  | |||  | |  |||  |   |||| || |  |||   |||  ||||   ||
Db         45 GATCAGTTTGTTTTCTGACTTGAAAGATATGAGATTGTTCATCAGGGTGGTGGTGGTGGT 104

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              |  ||||  |   |   |||  |||||  |  |  | ||  ||   | |||  | |  |
Db        105 GGTGATGATGATGATCATGCTCATAATGATGGTGATTAAAAGAAAGAAAATTTTATATCA 164

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTA 325
              |||  | |     | | |||  | ||  |||
Db        165 AATGTTACACTTCTGGCCCACAGAAAGACTA 195


RESULT 10
AK040558
LOCUS       AK040558               3362 bp    mRNA    linear   HTC 19-SEP-2003
DEFINITION  Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
            library, clone:A430106I07 product:phosphatidylinositol-4-phosphate
            5-kinase, type II, alpha, full insert sequence.
ACCESSION   AK040558
VERSION     AK040558.1  GI:26333816
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3362)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .3362
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A430106I07"
                     /db_xref="MGI:2405242"
                     /db_xref="taxon:10090"
                     /clone="A430106I07"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="0 day neonate"
     CDS             17. .1234
                     /note="unnamed protein product;
                     phosphatidylinositol-4-phosphate 5-kinase, type II, alpha
                     (MGD|MGI:1298206)
                     putative"
                     /codon_start=1
                     /protein_id="BAC30626.1"
                     /db_xref="GI:26333817"
                     /translation="MATPGNLGSSVLASKTKTKKKHFVAQKVKLFRASDPLLSVLMWG
                     VNHSINELSHVQIPVMLMPDDFKAYSKIKVDNHLFNKENMPSHFKFKEYCPMVFRNLR
                     ERFGIDDQDFQNSLTRSAPLPNDSQARSGARFHTSYDKRYVIKTITSEDVAEMHNILK
                     KYHQYIVECHGVTLLPQFLGMYRLNVDGVEIYVIVTRNVFSHRLSVYRKYDLKGSTVA
                     REASDKEKAKELPTLKDNDFINEGQKIYIDDNNKKIFLEKLKKDVEFLAQLKLMDYSL
                     LVGIHDVERAEQEEVECEENDGEEEGESDSTHPIGTPPDSPGNTLNSSPPLAPGEFDP
                     NIDVYAIKCHENAPRKEVYFMAIIDILTHYDAKKKAAHAAKTVKHGAGAEISTVNPEQ
                     YSKRFLDFIGHIL"
     polyA_signal    3338. .3343
                     /note="putative"
     polyA_site      3362
                     /note="putative"
ORIGIN

  Query Match             8.2%;  Score 37.4;  DB 11;  Length 3362;
  Best Local Similarity   53.0%;  Pred. No. 31;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        175 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGC 234
              ||||||  | |||  | |  |||  |   |||| || |  |||   |||  ||||   ||
Db       1655 GATCAGTTTGTTTTCTGACTTGAAAGATATGAGATTGTTCATCAGGGTGGTGGTGGTGGT 1714

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              |  ||||  |   |   |||  |||||  |  |  | ||  ||   | |||  | |  |
Db       1715 GGTGATGATGATGATCATGCTCATAATGATGGTGATTAAAAGAAAGAAAATTTTATATCA 1774

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTA 325
              |||  | |     | | |||  | ||  |||
```

```
Db       1775 AATGTTACACTTCTGGCCCACAGAAAGACTA 1805


RESULT 11
AU101070/c
LOCUS       AU101070                711 bp    mRNA    linear   EST 03-APR-2002
DEFINITION  AU101070 Rice panicle at flowering stage Oryza sativa (japonica
            cultivar-group) cDNA clone E0707, mRNA sequence.
ACCESSION   AU101070
VERSION     AU101070.1  GI:9865320
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from panicle at flowering stage (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="E0707"
                     /dev_stage="flowering stage"
                     /clone_lib="Rice panicle at flowering stage"
                     /note="Organ: panicle; Rice cDNA from panicle at flowering
                     stage"
ORIGIN

  Query Match             8.0%;  Score 36.6;  DB 9;  Length 711;
  Best Local Similarity   48.8%;  Pred. No. 27;
  Matches   99;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
              ||   | | ||  ||   |    |||| |  | | |       | |   | | ||| | | |
Db        458 GGTCACCGAAAGGGTCACCAAACGGATTGGAAGATCCAGCCTGAGGTGCAGGGTACTGAT 399

Qy        160 AACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 219
                |||||    |  || |       |  |   | |||||    |||     | | || ||
Db        398 GCTGCTGCTGCTGCTGTTGTTGGTATATCTGATTTTGTTGATAATGTGTCTGAAAATACA 339

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
               |||||||| | |  |  |||   ||  | | || || |  ||||| | |||.      |
Db        338 TGTTATGGTGATGTTGATGATCCTGGAAATGATGAGGCTGGTAGTATTGTTGCTGGTGTT 279

Qy        280 CAGAATGGTTTCAGAAATAATGC 302
              ||| ||||   |   ||  |||
Db        278 GAGACTGGTAGTACTGATGCTGC 256


RESULT 12
AZ180696
LOCUS       AZ180696                500 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_0181_A1_C06_SP6E Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library Strongylocentrotus purpuratus
            genomic clone Plate=181 Col=11 Row=E, genomic survey sequence.
ACCESSION   AZ180696
VERSION     AZ180696.1  GI:8353071
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
            Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H.
            and Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
   PUBMED   10920195
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 181  row: E  column: 11
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 500.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7668"
                     /clone="Plate=181 Col=11 Row=E"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
ORIGIN

  Query Match             7.8%;  Score 35.6;  DB 28;  Length 500;
  Best Local Similarity   50.9%;  Pred. No. 44;
  Matches   83;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        133 TTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGT 192
              ||   || | ||    | | |  || | | | ||    | |||| |    | |    |
Db          8 TTNTCCAATAATATCCATGATAACGATGATGATGATGATGATGATGATGATTGA 67

Qy        193 GTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGT 252
              | | |    | ||    |  |||  | | ||| | |   ||  ||||  |   | | |
Db         68 GACGATGATGACGATGATGATGATGATGATGATGAATACGATGATGATGATGATGAT 127

Qy        253 GCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
              |  ||| ||| || || |||| | | | |  ||| || ||||
Db        128 GATGATGATAATAATAATGAAATTAATAATGATGATTACAGAA 170


RESULT 13
CA141136/c
LOCUS       CA141136                664 bp    mRNA    linear   EST 24-SEP-2003
DEFINITION  SCJFRT2058D10.g RT2 Saccharum officinarum cDNA clone SCJFRT2058D10
            5', mRNA sequence.
ACCESSION   CA141136
VERSION     CA141136.1  GI:35034821
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
```

```
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 058  row: D  column: 10
            Seq primer: T7 Promoter Primer.
FEATURES              Location/Qualifiers
     source           1. .664
                      /organism="Saccharum officinarum"
                      /mol_type="mRNA"
                      /db_xref="taxon:4547"
                      /clone="SCJFRT2058D10"
                      /lab_host="DH10B"
                      /clone_lib="RT2"
                      /note="Organ: Root tips(0.3cm-long) from adult plants;
                      Vector: pSport1; Site_1: SalI; Site_2: NotI; An
                      unidirectional cDNA library generated from [Root
                      tips(0.3cm-long) from  adult plants]. cDNA was prepared
                      from polyA+ mRNA using SuperScript  Plasmid System Kit
                      (Invitrogen). The double-strand cDNAs were  fractionated
                      in a sepharose CL-2B 40cm-columns and fragments sizing
                      between 0.8 and 1.5 Kb were directionally cloned into the
                      vector. Details  of each source of RNA and library
                      construction can be obtained at
                      http://sucest.lad.ic.unicamp.br/public"
ORIGIN

  Query Match             7.8%;  Score 35.4;  DB 13;  Length 664;
  Best Local Similarity   46.2%;  Pred. No. 56;
  Matches  117;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy         22 GCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGC 81
              ||  |  ||      ||| || |   |    || |   | ||  | | ||  |||  | | |
Db        521 GCCTGCGAGTATGTCGGAGCTGAGGCACCTGGTTGCGTTGCAGACCTTGTACATGGCGAC 462

Qy         82 GGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATT 141
              |   |||     |   |   ||||||| |      ||||| || |   |   |||||
Db        461 GTGCGCGCGGGGCGCCATCCCGGCGGCGACCCCCACGGCCTGGCCAAGAACCTGAGCGCC 402

Qy        142 TATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAGCTGGTTACCCGTGTTGTTACC 201
                ||  |||| |||| |  |  |||      ||  |     ||  || | ||  |   |
Db        401 CTGCACGACGGCTCCGGTGGCCGTGCTGGCGGTGTGCGTGCCGTGGCCGGCGTCATCGAC 342

Qy        202 CATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAAT 261
              |    |  | |  ||| |     |||   | ||   ||| |||   | |  | ||| |
Db        341 CGGCGACGAGGAGTTGCTGCCGGCGGTGGGGACGGACATGAAGGAGCGAGCTCCGATGAG 282

Qy        262 AGTACTATTGAAC 274
                |  | ||| ||
Db        281 CTTGTTGTTGCAC 269


RESULT 14
AQ660801
LOCUS       AQ660801                 556 bp    DNA     linear   GSS 23-JUN-1999
DEFINITION  Sheared DNA-28K9.TF Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-28K9, genomic survey sequence.
ACCESSION   AQ660801
VERSION     AQ660801.1  GI:5168569
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-28K9.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .556
                      /organism="Trypanosoma brucei"
                      /mol_type="genomic DNA"
                      /strain="TREU927/4 GUTat 10.1"
                      /db_xref="taxon:5691"
                      /clone="Sheared DNA-28K9"
                      /clone_lib="Sheared DNA"
                      /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                      Institute for Genomic Research (TIGR), Rockville, MD.
                      Genomic DNA isolated from a cloned population of
                      Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                      sheared to give a tight size distribution (approx 2 kb).
                      The v + i method used for the library construction is
                      described in detail in Smith, H.O. and Venter, J.C.
                      (Making small insert libraries for whole genome shotgun
                      sequencing projects.  In Genome Sequencing: A Practical
                      Approach, eds. M. Vaudin and B. Barell, Oxford University
                      Press, 1999)."
ORIGIN

  Query Match             7.7%;  Score 35;  DB 28;  Length 556;
  Best Local Similarity   47.1%;  Pred. No. 67;.
  Matches  107;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        185 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAG 244
              |||| || |   ||||  ||||   |||  | |    |    | |||   ||  | ||
Db        237 TTACACGCGAACTTACGGATGAGCGGGCGAAGGTCTCTGGCCTTCACGAAACCCTCGCAG 296

Qy        245 GCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCA 304
              | | |||||||            ||  ||  ||  |  |  |  ||  |  |   || |
Db        297 AACGGAGTGCGGAACTGGAACTAATGAAAGAGAAACTTAGCGATGCCACAGTTGTCGCAA 356

Qy        305 CCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATA 364
              |   ||||    ||||  | |  |||| | | |         | |  ||  |  | | |||
Db        357 CTGACGACACCGGGAAGCCGATCAACTACAAGAGGGAGTGTGACGGATTGGCTGCTGATA 416

Qy        365 ACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTG 411
              |   |   || || ||  ||  ||     | || ||  | | |||
Db        417 ATAACCGACTTGTGAAGGAGTACGAGCGCGTTTGCACGGAACTGGAG 463


RESULT 15
AQ952353
LOCUS       AQ952353                 679 bp    DNA     linear   GSS 27-JAN-2000
DEFINITION  Sheared DNA-32L14.TF Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-32L14, genomic survey sequence.
ACCESSION   AQ952353
VERSION     AQ952353.1  GI:6775618
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
```

```
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-32L14.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .679
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-32L14"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
ORIGIN

  Query Match          7.7%;  Score 35;  DB 28;  Length 679;
  Best Local Similarity  47.1%;  Pred. No. 73;
  Matches  107;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        185 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAG 244
              |||| || |   ||||  ||||   |||  | |     |      | |||   ||   | ||
Db        212 TTACACGCGAACTTACGGATGAGCGGGCGAAGGTCTCTGGCCTTCACGAAACCCTCGCAG 271

Qy        245 GCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCA 304
                | | |||||||          ||  ||  ||  |  |  |  |  || |  |   || |
Db        272 AACGGAGTGCGGAACTGGAACTAATGAAAGAGAAACTTAGCGATGCCACAGTTGTCGCAA 331

Qy        305 CCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATA 364
              |   ||||    ||||  | |  |||| | | |       | |  ||  |  | | |||
Db        332 CTGACGACACCGGGAAGCCGATCAACTACAAGAGGGAGTGTGACGGATTGGCTGCTGATA 391

Qy        365 ACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTG 411
              |   |   || || ||  ||  ||      | || ||  | | ||| |
Db        392 ATAACCGACTTGTGAAGGAGTACGAGCGCGTTTGCACGGAACTGGAG 438


Search completed: March 15, 2004, 10:09:28
Job time : 1967.4 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 19:02:18 ; Search time 2269.8 Seconds
(without alignments)
8707.562 Million cell updates/sec

Title: US-09-543-407-27
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		401.6	88.1	456	6	I44909	I44909 Sequence 58
2		401.6	88.1	2067	1	SEU43280	U43280 Salmonella
3		393.6	86.3	5103	1	STAJ2301	AJ002301 Salmonell
4		393.6	86.3	22411	1	AE008749	AE008749 Salmonell
5		392	86.0	254050	1	AL627269	AL627269 Salmonell
6	c	392	86.0	301983	1	AE016840	AE016840 Salmonell
7		386.2	84.7	1048	1	STAGFBA	AJ000514 Salmonell
8		277.4	60.8	361	6	I44908	I44908 Sequence 56
9		272	59.6	2889	1	CSP515700	AJ515700 Citrobact
10		230.4	50.5	4680	1	ECCSGABDG	X90754 E.coli csgG
11		230.4	50.5	10346	1	AE000205	AE000205 Escherich
12		230.4	50.5	15047	1	D90741	D90741 Escherichia
13		228.8	50.2	456	6	AX814811	AX814811 Sequence
14		228.8	50.2	648	1	ECOCSGAA	L04979 Escherichia
15		224.4	49.2	2920	1	CFR515701	AJ515701 Citrobact
16		222.2	48.7	1711	1	AF275733	AF275733 Escherich
17		222.2	48.7	10190	1	AE005315	AE005315 Escherich
18		222.2	48.7	327773	1	AP002554	AP002554 Escherich
19		211	46.3	306358	1	AE016759	AE016759 Escherich
20		190.4	41.8	2883	1	ESA515702	AJ515702 Enterobac
21		174.6	38.3	230	1	SEU53207	U53207 Salmonella
22		164.8	36.1	10370	1	AE015131	AE015131 Shigella
23		164.8	36.1	292504	1	AE016981	AE016981 Shigella
24	c	141.8	31.1	437	1	AF237726	AF237726 Shigella
25		87	19.1	19201	1	D90742	D90742 Escherichia
26		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27		48.2	10.6	78	6	AX814809	AX814809 Sequence
28	c	46.6	10.2	2000	6	AX655393	AX655393 Sequence
29		44.8	9.8	72	6	AX814798	AX814798 Sequence
30		37.4	8.2	2560	3	AY113223	AY113223 Drosophil
31	c	36.8	8.1	62909	6	AR408753	AR408753 Sequence
32	c	36.8	8.1	62909	6	AX067457	AX067457 Sequence
33		36.4	8.0	334108	1	AE016857	AE016857 Pseudomon
34	c	36.2	7.9	1788	8	AF105426	AF105426 Cynodon d
35		35.8	7.9	1602	6	AX482643	AX482643 Sequence
36		35.8	7.9	2101	3	AY052147	AY052147 Drosophil
37		35.8	7.9	48042	2	AC020217	AC020217 Drosophil
38	c	35.8	7.9	169705	3	AC104704	AC104704 Drosophil
39		35.8	7.9	171907	2	AC116538	AC116538 Drosophil
40	c	35.8	7.9	279904	3	AE003527	AE003527 Drosophil
41	c	35.2	7.7	120938	8	AC007017	AC007017 Arabidops
42		34.8	7.6	346301	1	BX640432	BX640432 Bordetell
43	c	34.6	7.6	11279	1	AE015613	AE015613 Shewanell
44		34.6	7.6	122939	8	AC116369	AC116369 Oryza sat
45		34.4	7.5	561	5	AF257332	AF257332 Xenopus l

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             88.1%;  Score 401.6;  DB 6;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 4.7e-104;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              |  |||  |      |  |  ||  |  |      |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match             88.1%;  Score 401.6;  DB 1;  Length 2067;
  Best Local Similarity   92.5%;  Pred. No. 5.2e-104;
```

```
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              |  |||   |       |  |  ||   |     |       |||  ||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301                5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNPAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN
```

Query Match 86.3%; Score 393.6; DB 1; Length 5103;
Best Local Similarity 91.4%; Pred. No. 1.1e-101;
Matches 417; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              |  |||  |  |  ||  |  |  |||  ||||||||  |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421
```

```
RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA

COMMENT COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES Location/Qualifiers

source 1. .22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

gene 434. .1308
/gene="phoH"
/note="synonym: STM1126"

RBS 434. .439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"

CDS 454. .1308
/gene="phoH"
/note="similar to E. coli PhoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=11
/product="PhoB-dependent ATP-binding pho regulon component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
QRTLHAYS"

gene complement(1414. .2303)
/gene="STM1127"

CDS complement(1414. .2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29% identity in aa 15 - 264"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
QARESKLRTMNALDMTK"

RBS complement(2298. .2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"

gene complement(2580. .4088)
/gene="STM1128"

CDS complement(2580. .4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22% identity in aa 7 - 478"
/codon_start=1
/transl_table=11
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"

RBS complement(4083. .4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"

gene complement(4413. .5099)
/gene="STM1129"

CDS complement(4413. .5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 - 226"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
WFSHAVKR"

RBS complement(5094. .5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"

gene 5599. .6759
/gene="STM1130"

CDS 5599. .6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41% identity in aa 40 - 403"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"

gene 6793. .7497
/gene="STM1131"

RBS 6793. .6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match             86.3%;  Score 393.6;  DB 1;  Length 22411;
  Best Local Similarity   91.4%;  Pred. No. 1.2e-101;
  Matches  417;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                | |||  |          |  ||   |    |      |||  |||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269              254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

/note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature"
gene 3001. .4011
/gene="STY1079"
/note="synonym: pyrD"
CDS 3001. .4011
/gene="STY1079"
/note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q8Z7S9"
/db_xref="SWISS-PROT:Q8Z7S9"
/translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
HI"
misc_feature 3136. .3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHOdehase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
misc_feature 3238. .3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
misc_feature 3874. .3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
gene 4174. .4719
/gene="STY1080"
CDS 4174. .4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepency in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q8Z7S8"
/translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
PCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
IMNDRLKPQVHCHSFSLEQAV"
gene complement(4716. .5825)
/gene="STY1081"
CDS complement(4716. .5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Fasta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q8Z7S7"
/db_xref="SPTREMBL:Q8Z7S7"
/translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
KHPSGEPLATLQAPRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
misc_feature complement(4737. .4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
gene 5924. .8032
/gene="STY1082"
CDS 5924. .8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q8Z7S6"
/db_xref="SPTREMBL:Q8Z7S6"
/translation="MNSLPASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
LSPDFARNRQIHNCWLIRAA"
misc_feature 5927. .7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, score 781.30, E-value 3.7e-231"
misc_feature 6506. .6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
gene 8045. .9952
/gene="STY1083"
CDS 8045. .9952
/gene="STY1083"
/note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa overlap"

Query Match 86.0%; Score 392; DB 1; Length 254050;
Best Local Similarity 91.2%; Pred. No. 4e-101;
Matches 416; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```
Qy        1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy       61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy      121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
            |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy      181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy      241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              |  |||  |  |  ||  |  |  |  |||  |||||||| |||
Db    89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae;. Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIEREL PTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.0%;  Score 392;  DB 1;  Length 301983;
Best Local Similarity   91.2%;  Pred. No. 4e-101;
  Matches  416;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              |  |||  |  |  ||  |  |  |  |  |||  |||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             84.7%;  Score 386.2;  DB 1;  Length 1048;
  Best Local Similarity   90.5%;  Pred. No. 1.3e-99;
  Matches  412;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                | |||  |          |  ||   |      |||  ||||||||  |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  ||||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             60.8%;  Score 277.4;  DB 6;  Length 361;
  Best Local Similarity   89.3%;  Pred. No. 2e-68;
  Matches  299;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTAT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCC 303
              | |||  |          |  ||   |      |||  ||||||||  |||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           59.6%;  Score 272;  DB 1;  Length 2889;
  Best Local Similarity  76.3%;  Pred. No. 8.1e-67;
  Matches  348;  Conservative    0;  Mismatches  105;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |     ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||    |  |||||| ||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||    ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                |  |||                 | | ||   |    | | |  |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || |||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               |||||||||| ||||| || ||||| || || |||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
```

```
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene                 complement(1434. .1823)
                     /gene="csgE"
CDS                  complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
gene                 complement(1828. .2478)
                     /gene="csgD"
CDS                  complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
misc_feature         complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
misc_feature         3140
                     /note="transcriptional start site for csgBA operon"
gene                 3233. .3688
                     /gene="csgB"
CDS                  3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
gene                 3729. .4184
                     /gene="csgA"
CDS                  3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
sig_peptide          3729. .3788
                     /gene="csgA"
gene                 4243. .4575
                     /gene="orfC"
CDS                  4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           50.5%;  Score 230.4;  DB 1;  Length 4680;
  Best Local Similarity 69.1%;  Pred. No. 7.1e-55;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||     || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                || ||||          | | ||   |     || ||  ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 11
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
```

```
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES              Location/Qualifiers
     source           1. .10346
                      /organism="Escherichia coli K12"
                      /mol_type="genomic DNA"
                      /strain="K12"
                      /sub_strain="MG1655"
                      /db_xref="taxon:83333"
     repeat_region    complement(<1. .48)
                      /note="IS3"
     promoter         <1. .5
                      /note="factor Sigma70; predicted +1 start at 1094689"
     promoter         5. .32
                      /note="factor Sigma70; predicted +1 start at 1094716"
     gene             69. .392
                      /gene="b1028"
     CDS              69. .392
                      /gene="b1028"
                      /function="orf; Unknown"
                      /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                      to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                      SW: P43098"
                      /codon_start=1
                      /transl_table=11
                      /product="orf, hypothetical protein"
                      /protein_id="AAC74113.1"
                      /db_xref="GI:1787266"
                      /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                      LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                      AKVFK"
     promoter         275. .299
                      /gene="b1028"
                      /note="factor Sigma32; predicted +1 start at 1094983"
     gene             389. .1375
                      /gene="ycdU"
                      /note="synonym: b1029"
     CDS              389. .1375
                      /gene="ycdU"
                      /function="orf; Unknown"
                      /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                      to 191 residues of an approx. 288 aa protein Y320_MYCGE
                      SW: P47562"
                      /codon_start=1
                      /transl_table=11
                      /product="orf, hypothetical protein"
                      /protein_id="AAC74114.1"
                      /db_xref="GI:1787267"
                      /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                      GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                      LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                      SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                      FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                      ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind     1253. .1280
                      /gene="ycdU"
                      /note="central position to predicted promoter: -104"
                      /bound_moiety="DeoR predicted site"
     protein_bind     1253. .1280
                      /gene="ycdU"
                      /note="No predicted promoter"
                      /bound_moiety="DeoR predicted site"
     promoter         1337. .1365
                      /gene="ycdU"
                      /note="factor Sigma70; predicted +1 start at 1096049"
     gene             1494. .1745
                      /gene="b1030"
     CDS              1494. .1745
                      /gene="b1030"
                      /function="orf; Unknown"
                      /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                      to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                      P49065"
                      /codon_start=1
                      /transl_table=11
                      /product="orf, hypothetical protein"
                      /protein_id="AAC74115.1"
                      /db_xref="GI:1787268"
                      /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                      VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene             1513. .1926
                      /gene="ycdV"
                      /note="synonym: b1031"
     CDS              1513. .1926
                      /gene="ycdV"
                      /function="putative structure; Not classified"
                      /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                      to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                      P05759"
                      /codon_start=1
                      /transl_table=11
                      /product="putative ribosomal protein"
                      /protein_id="AAC74116.1"
                      /db_xref="GI:1787269"
                      /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                      GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                      SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene             complement(2111. .2198)
                      /gene="serX"
                      /note="synonym: b1032"
     tRNA             complement(2111. .2198)
                      /gene="serX"
                      /product="tRNA-Ser"
                      /note="anticodon: GGA; CG Site No. 17878"
                      /function="RNA; tRNA"
     promoter         2269. .2296
```

```
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

  Query Match           50.5%;  Score 230.4;  DB 1;  Length 10346;
  Best Local Similarity   69.1%;  Pred. No. 7.5e-55;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| ||||||||| || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              || ||||                 | | ||   |   || ||  ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448


RESULT 12
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                   Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .15047
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:83333"
                     /map="23.7-24.0 min"
                     /clone="Kohara clone #231"
                     /note="Nucleotide position 1097101-1112147 from the
```

initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIPYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS 2906. .3460
/note="ORF_ID:o231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=11
/product="Hypothetical protein HI1543"
/protein_id="BAA35825.1"
/db_xref="GI:4062608"
/translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
RTMAPLTRDAISAMWDELEEDSEE"
CDS 3514. .4053
/note="unnamed protein product; ORF_ID:o231#4"
/codon_start=1
/transl_table=11
/protein_id="BAA35826.1"
/db_xref="GI:4062609"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
LAARSAKTAHREQEIKNKA"
gene complement(4117. .4950)
/gene="csgG"
CDS complement(4117. .4950)
/gene="csgG"
/note="ORF_ID:o231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component CsgG precursor."
/protein_id="BAA35827.1"
/db_xref="GI:4062610"
/translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
S"
gene complement(4977. .5393)
/gene="csgF"
CDS complement(4977. .5393)
/gene="csgF"
/note="ORF_ID:o231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component CsgF precursor."
/protein_id="BAA35828.1"
/db_xref="GI:4062611"
/translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene complement(5418. .5807)
/gene="csgE"
CDS complement(5418. .5807)
/gene="csgE"
/note="ORF_ID:o231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component CsgE precursor."
/protein_id="BAA35829.1"
/db_xref="GI:4062612"
/translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
IQTEEALNRRQINQALLSTGDLAHDEF"
gene complement(5812. .6462)
/gene="csgD"
CDS complement(5812. .6462)
/gene="csgD"
/note="ORF_ID:o231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=11
/product="Probable csgAB operon transcriptional regulatory protein."
/protein_id="BAA35830.1"
/db_xref="GI:4062613"
/translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene 7217. .7672
/gene="csgB"
CDS 7217. .7672

Query Match 50.5%; Score 230.4; DB 1; Length 15047;
Best Local Similarity 69.1%; Pred. No. 7.7e-55;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | ||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952
```

```
Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              || ||||                 | | ||   |   || ||  ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  ||| |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168
```

RESULT 13
AX814811
LOCUS AX814811 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
TITLE Peptides
JOURNAL Patent: WO 03064446-A 15 07-AUG-2003;
Hansa Medical Research Aktiebolag (SE)
FEATURES Location/Qualifiers
source 1. .456
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
CDS 1. .456
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAE83541.1"
/db_xref="GI:39104002"
/translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGN

```
  Query Match             50.2%;  Score 228.8;  DB 6;  Length 456;
  Best Local Similarity   68.9%;  Pred. No. 1.8e-54;
  Matches  314;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||| | || || || |||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| ||||||||   || |||| |||| |||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||    ||     ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              || ||||                 | | ||   |   || ||  ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  ||| |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

RESULT 14
ECOCSGAA
LOCUS ECOCSGAA 648 bp DNA linear BCT 13-JUL-1993
DEFINITION Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION L04979
VERSION L04979.1 GI:290424
KEYWORDS csgA gene; curli organelle; fibronectin-binding protein.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 648)
AUTHORS Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
TITLE The RpoS sigma factor relieves H-NS-mediated transcriptional
repression of csgA, the subunit gene of fibronectin-binding curli
in Escherichia coli
JOURNAL Mol. Microbiol. 7 (4), 523-536 (1993)
MEDLINE 93211294
PUBMED 8459772
COMMENT On Jun 11, 1993 this sequence version replaced gi:145630.
Original source text: Escherichia coli (sub_strain W3110, strain
K-12) (library: Kohara) DNA.
FEATURES Location/Qualifiers
source 1. .648
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
/tissue_lib="Kohara"
gene 83. .538
/gene="csgA"
CDS 83. .538
/gene="csgA"
/note="precursor protein"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="curlin subunit"
/protein_id="AAA23616.1"
/db_xref="GI:290425"
/translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
sig_peptide 83. .142
/gene="csgA"
/evidence=experimental
mat_peptide 143. .535
/gene="csgA"
/product="curlin subunit"
/evidence=experimental
ORIGIN

```
  Query Match             50.2%;  Score 228.8;  DB 1;  Length 648;
  Best Local Similarity   68.9%;  Pred. No. 1.8e-54;
  Matches  314;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||| | || || || |||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142
```

```
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||     || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                || ||||                    | | ||   |    || ||  ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

```
RESULT 15
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN

  Query Match             49.2%;  Score 224.4;  DB 1;  Length 2920;
  Best Local Similarity   70.0%;  Pred. No. 3.7e-53;
  Matches  319;  Conservative    0;  Mismatches  131;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||    ||| ||||  ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||    ||||| || ||   |||   ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              ||  |||  | |        |  |        |         |  |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| ||||| |||||||| || |||||  | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||| |||||||||||||| || ||||| ||||  ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

```
Search completed: March 15, 2004, 07:24:15
Job time : 2274.2 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 18:52:53 ; Search time 290.4 Seconds
(without alignments)
6670.724 Million cell updates/sec

Title: US-09-543-407-27
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	3	AAC64630	Aac64630 AgfA::PT3
2	401.6	88.1	456	2	AAQ87467	Aaq87467 AgfA sequ
3	401.6	88.1	456	2	AAT74142	Aat74142 Salmonell
4	401.6	88.1	456	3	AAC64617	Aac64617 Salmonell
5	376	82.5	456	3	AAC64627	Aac64627 AgfA::PT3
6	360	78.9	456	3	AAC64626	Aac64626 AgfA::PT3
7	352	77.2	456	3	AAC64625	Aac64625 AgfA::PT3
8	350.4	76.8	456	3	AAC64628	Aac64628 AgfA::PT3
9	350.4	76.8	456	3	AAC64622	Aac64622 AgfA::PT3
10	350.4	76.8	456	3	AAC64629	Aac64629 AgfA::PT3
11	350.4	76.8	456	3	AAC64623	Aac64623 AgfA::PT3
12	348.8	76.5	456	3	AAC64624	Aac64624 AgfA::PT3
13	344	75.4	456	3	AAC64631	Aac64631 AgfA::PT3
14	277.4	60.8	361	2	AAQ73066	Aaq73066 AgfA sequ
15	277.4	60.8	361	2	AAT74141	Aat74141 Salmonell
16	228.8	50.2	456	3	AAC64619	Aac64619 Escherich
17	228.8	50.2	456	9	ACF36153	Acf36153 E. coli C
18	218	47.8	646	2	AAQ62647	Aaq62647 Fibronect
19	154.4	33.9	369	2	AAQ62646	Aaq62646 FNB curli
20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
c 21	49.8	10.9	78	3	AAC64609	Aac64609 AgfA (SEF
22	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
23	48	10.5	48	3	AAC64621	Aac64621 Leishmani
24	48	10.5	78	3	AAC64610	Aac64610 AgfA (SEF
c 25	46.6	10.2	2000	7	ADA71938	Ada71938 Rice gene
26	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
27	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
28	41	9.0	100	7	ACD68808	Acd68808 E. coli K
29	40.6	8.9	78	3	AAC64606	Aac64606 SefA (SEF
c 30	40.6	8.9	78	3	AAC64605	Aac64605 SefA (SEF
c 31	36.8	8.1	5838	7	ACA38893	Aca38893 Prokaryot
c 32	36.8	8.1	62909	4	AAF28545	Aaf28545 Genomic f
33	35.8	7.9	1602	6	ABS51383	Abs51383 cDNA enco
34	35.8	7.9	2030	4	ABL21031	Abl21031 Drosophil
35	35.8	7.9	2077	4	ABL21033	Abl21033 Drosophil
36	35.8	7.9	2122	4	ABL07821	Abl07821 Drosophil
37	35.8	7.9	5421	4	ABL21032	Abl21032 Drosophil
38	35.8	7.9	6365	4	ABL21030	Abl21030 Drosophil
39	35.8	7.9	6371	4	ABL07820	Abl07820 Drosophil
40	35.8	7.9	8661	4	ABL16184	Abl16184 Drosophil
41	34.6	7.6	1355	7	ABZ37481	Abz37481 M13-III n
42	34.4	7.5	1138	4	ABL05727	Abl05727 Drosophil
43	34.4	7.5	1145	4	ABL19547	Abl19547 Drosophil
44	34.4	7.5	1281	4	ABL06571	Abl06571 Drosophil
c 45	34.4	7.5	65589	6	ABA90521_23	Continuation (24 o

ALIGNMENTS

RESULT 1
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.1e-139;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 2; Length 456;
Best Local Similarity 92.5%; Pred. No. 9.3e-122;
Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              | |||  |      | | ||  |   |     ||| ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.1%;  Score 401.6;  DB 2;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 9.3e-122;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              |  |||  |      |  |  ||  |   |      |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 3; Length 456;
Best Local Similarity 92.5%; Pred. No. 9.3e-122;
Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                |  |||  |      |  |  ||  |   |     |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36351.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 82.5%; Score 376; DB 3; Length 456;
Best Local Similarity 89.0%; Pred. No. 2.5e-113;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                |  |||  |      |  || |   | |  |       ||    |   | ||
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match           78.9%;  Score 360;  DB 3;  Length 456;
  Best Local Similarity  86.8%;  Pred. No. 4.7e-108;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| |||   |   ||||||   | |   | |   ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                | |||   |       | |  ||   |     |      ||| ||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             77.2%;  Score 352;  DB 3;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.1e-105;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||      |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                | |||  |      |  |  ||   |    |      |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 6.9e-105;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 6.9e-105;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 6.9e-105;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                |  |||    |        |  |  ||    |      |       |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 6.9e-105;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                | |||  |     | | ||  |   |      |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||   | ||||  ||      |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64624
ID   AAC64624 standard; DNA; 456 BP.
XX
AC   AAC64624;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36348.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

  Query Match            76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 2.3e-104;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||  ||    | | | |  |    | |     ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                | |||  |     | | ||  |   |      |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             75.4%;  Score 344;  DB 3;  Length 456;
  Best Local Similarity   84.6%;  Pred. No. 8.9e-103;
  Matches  386;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              |  |||   |     |  |  ||  |   |     |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |   |              |   | |. ||   ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             60.8%;  Score 277.4;  DB 2;  Length 361;
  Best Local Similarity   89.3%;  Pred. No. 7e-81;
  Matches  299;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTAT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCC 303
              |  |||   |      |  |  ||   |    |      |||  ||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             60.8%;  Score 277.4;  DB 2;  Length 361;
  Best Local Similarity   89.3%;  Pred. No. 7e-81;
  Matches  299;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTAT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCC 303
              |  |||   |      |  |  ||   |    |      |||  ||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 15, 2004, 04:14:45
```

Job time : 292.4 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:01:03 ; Search time 56.8 Seconds
(without alignments)
4455.240 Million cell updates/sec

Title: US-09-543-407-27
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	401.6	88.1	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	277.4	60.8	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	36.8	8.1	62909	4	US-09-596-002-32	Sequence 32, Appl
c	4	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	5	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
c	6	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	7	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	8	30.2	6.6	1200	4	US-09-134-000C-1704	Sequence 1704, Ap
	9	30.2	6.6	2226	4	US-09-489-039A-683	Sequence 683, App
c	10	30.2	6.6	536165	4	US-09-214-808-1	Sequence 1, Appli
c	11	30	6.6	2436	4	US-09-540-236-492	Sequence 492, App
c	12	30	6.6	65792	4	US-09-596-002-31	Sequence 31, Appl
	13	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	14	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	15	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
	16	29.8	6.5	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	17	29.6	6.5	2426	4	US-09-023-655-1156	Sequence 1156, Ap
c	18	29.6	6.5	2557	2	US-09-044-506A-1	Sequence 1, Appli
c	19	29.6	6.5	2557	3	US-08-478-208-31	Sequence 31, Appl
c	20	29.6	6.5	2557	4	US-09-336-536-72	Sequence 72, Appl
c	21	29.6	6.5	2557	6	5264554-1	Patent No. 5264554
c	22	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	23	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	24	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	25	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	26	29.4	6.4	2040	4	US-09-540-236-232	Sequence 232, App
c	27	29.4	6.4	39003	4	US-09-596-002-21	Sequence 21, Appl
	28	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
	29	29.2	6.4	564	4	US-09-134-001C-29	Sequence 29, Appl
c	30	29.2	6.4	729	4	US-09-134-001C-31	Sequence 31, Appl
c	31	29.2	6.4	1540	4	US-09-149-476-151	Sequence 151, App
c	32	29.2	6.4	2052	1	US-08-443-104-5	Sequence 5, Appli
c	33	29.2	6.4	2052	1	US-08-238-130-6	Sequence 6, Appli
c	34	29.2	6.4	2052	1	US-08-442-859-5	Sequence 5, Appli
c	35	29.2	6.4	2052	2	US-08-398-489-5	Sequence 5, Appli
c	36	29.2	6.4	2052	2	US-08-894-772-1	Sequence 1, Appli
c	37	29.2	6.4	2052	2	US-09-207-844-1	Sequence 1, Appli
c	38	29.2	6.4	2052	4	US-09-252-509-1	Sequence 1, Appli
c	39	29.2	6.4	2052	5	PCT-US95-05534-5	Sequence 5, Appli
	40	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	41	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	42	29	6.4	448	4	US-09-621-976-13374	Sequence 13374, A
c	43	29	6.4	534	4	US-09-489-039A-2195	Sequence 2195, Ap
	44	29	6.4	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
	45	29	6.4	6877	1	US-08-347-340-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 88.1%; Score 401.6; DB 1; Length 456;
Best Local Similarity 92.5%; Pred. No. 2.8e-131;
Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              | |||  |    |  | ||  |   |     ||| ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             60.8%;  Score 277.4;  DB 1;  Length 361;
  Best Local Similarity   89.3%;  Pred. No. 1.2e-87;
  Matches  299;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |||||  ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTAT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCC 303
              | |||  |    |  | ||  |   |     ||| ||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

```
RESULT 3
US-09-596-002-32/c
; Sequence 32, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:      Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:       PERL Program
; SEQ ID NO 32
;   LENGTH: 62909
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte template ID No. 6632636 32
;  PUBLICATION INFORMATION:
US-09-596-002-32

  Query Match              8.1%;  Score 36.8;  DB 4;  Length 62909;
  Best Local Similarity   50.0%;  Pred. No. 0.093;
  Matches   92;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTATGATCAGCTGGTTAC 257
              ||  |  || ||||        |||| || | | | | | || ||| | |  ||  ||||
Db      43939 TGGTAGTACTATTATTATTTACGGTGATTGCACCTGTGGCGGTTAGGTTTTGCGTGTTAA 43880

Qy        258 CCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCCACCATCGACCAGTG 317
```

```
                | | |   | |  | ||||          || |    | |  |||||| |
Db      43879 CTGCTTCAAGATCAGTTAAATTTTTAGCAAGTTTAACTTTCAGACCACCACCATTATCAT 43820

Qy        318 GAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGT 377
              | |  | |  || ||||| ||| |||||| ||| |  ||| | |  | |  |      |
Db      43819 TATCTTTTACCACACCGATGTTATTGTCGGTTAATTCACCGGCATTTGCTCGCCCGCCTT 43760

Qy        378 TAAT 381
              ||||
Db      43759 TAAT 43756


RESULT 4
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
;  GENERAL INFORMATION:
;    APPLICANT:  Kim, Jihyun Francis
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;    TITLE OF INVENTION:  ERWINIA AMYLOVORA AND ITS USE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,927
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,108
;      FILING DATE:  06-AUG-1977
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1581
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1344 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.1;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;  LENGTH: 1344
;  TYPE: DNA
;  ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.1;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 6
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: PLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;  LENGTH: 4403765
;  TYPE: DNA
;  ORGANISM: Mycobacterium tuberculosis
;  FEATURE:
;  OTHER INFORMATION: CDC 1551
;  OTHER INFORMATION: "n" bases at various positions throughout the sequence
;  OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4403765;
  Best Local Similarity   51.9%;  Pred. No. 87;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || ||||||| || ||  |  ||  | |||||| |||| |  |   || ||
Db    1357218 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357159
```

```
Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
              ||     | | |    ||  | ||| | ||   |   ||  || || | |  |||
Db    1357158 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357099

Qy        165 TGCGCTTGCTCTG 177
              ||   |  |||
Db    1357098 CCCGGACGTGCTG 1357086
```

RESULT 7
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 6.7%; Score 30.6; DB 3; Length 4411529;
Best Local Similarity 51.9%; Pred. No. 87;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || |||||||| || ||   |  ||   | ||||| |||| |  |   || ||
Db    1357750 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357691

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
              ||     | | |    ||  | ||| | ||   |   ||  || || | |  |||
Db    1357690 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357631

Qy        165 TGCGCTTGCTCTG 177
              ||   |  |||
Db    1357630 CCCGGACGTGCTG 1357618
```

RESULT 8
US-09-134-000C-1704
; Sequence 1704, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1704
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1704

Query Match 6.6%; Score 30.2; DB 4; Length 1200;
Best Local Similarity 55.1%; Pred. No. 2.1;
Matches 59; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy        266 TTACCCATGAAATGGCACATGCATTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              | ||||||||||  |     |   || | | |||||||||   | || || | | |    |
Db        509 TAACCCATGAAAAAGATGGTCTTTTAAAAGATAATGCCATGTTAGATCAATTGTATCAAA 568

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCG 372
              ||||   |||  |  ||  | |  | ||  |  || |  ||| | |
Db        569 AAAATAAAGATTATTTTGATGCCTTAAACATCTATATTGCCGCTGGG 615
```

RESULT 9
US-09-489-039A-683
; Sequence 683, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 683
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-683

Query Match 6.6%; Score 30.2; DB 4; Length 2226;
Best Local Similarity 60.2%; Pred. No. 3;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy        211 ACCCAGAGCGGTTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACC 270
              | | |||||||||| ||  ||| ||  |||| ||||| ||| ||  |    | || | |
Db        523 AGCAAGAGCGGTTTTGCGAACAGCAACGATCATGATGAGCGCGTCGCGGCGGCCGTCAGC 582

Qy        271 CATGAAATGGCACATGCATTCAG 293
                | ||  |  || |||| | |
Db        583 GGCGGAACAGACCACGCATCCGG 605
```

RESULT 10
US-09-214-808-1/c
; Sequence 1, Application US/09214808A
; Patent No. 6475793
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; Patent No. 6475793
; TITLE OF INVENTION: Plasmid
; FILE REFERENCE: CARP0068
; CURRENT APPLICATION NUMBER: US/09/214,808A
; CURRENT FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: PCT/IB97/00950
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-214-808-1

Query Match 6.6%; Score 30.2; DB 4; Length 536165;
Best Local Similarity 58.2%; Pred. No. 63;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```
Qy        231 CGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATT 290
              || |||||   |||||| ||| |     |   | | ||||    |||| ||   ||  | |
Db     311150 CGCCGCCGCCGATGATCGGCTCGCCGGCGCGGATCTTACTATGGAAACGGTCGATCAACT 311091

Qy        291 CAGAAATAATGCCACCATCGACCAGTGGAAC 321
              |||||  | ||  ||  | ||| || | | |
Db     311090 CAGAACGAGTGGGACGCTTGACAAGAGTATC 311060


RESULT 11
US-09-540-236-492/c
; Sequence 492, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2005-001
;  CURRENT APPLICATION NUMBER: US/09/540,236
;  CURRENT FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 492
;   LENGTH: 2436
;   TYPE: DNA
;   ORGANISM: M.catarrhalis
US-09-540-236-492

  Query Match             6.6%;  Score 30;  DB 4;  Length 2436;
  Best Local Similarity   57.4%;  Pred. No. 3.7;
  Matches   54;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        276 AATGGCACATGCATTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              |||   | |||| |||| |||||| ||||  |||| ||  ||  |     ||       |
Db       1441 AATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAAACTGTCACCT 1382

Qy        336 TATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              || ||  ||  |||| ||  ||||   |  |||
Db       1381 CATGACCATCAACCAAAACTTCGGTTGCATTGCC 1348


RESULT 12
US-09-596-002-31/c
; Sequence 31, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:      Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:      PERL Program
; SEQ ID NO 31
;   LENGTH: 65792
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte template ID No. 6632636 31
;  PUBLICATION INFORMATION:
US-09-596-002-31

  Query Match             6.6%;  Score 30;  DB 4;  Length 65792;
  Best Local Similarity   57.4%;  Pred. No. 23;
  Matches   54;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        276 AATGGCACATGCATTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              |||   | |||| |||| |||||| ||||  |||| ||  ||  |     ||       |
Db       6749 AATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAAACTGTCACCT 6690

Qy        336 TATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              || ||  ||  |||| ||  ||||   |  |||
Db       6689 CATGACCATCAACCAAAACTTCGGTTGCATTGCC 6656


RESULT 13
US-09-252-991A-10904
; Sequence 10904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10904
;   LENGTH: 1008
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904

  Query Match             6.5%;  Score 29.8;  DB 4;  Length 1008;
  Best Local Similarity   54.0%;  Pred. No. 2.6;
  Matches   61;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
              | | || | |  |  || | |  | ||| |  |    ||||  ||||||  | |     |
Db        141 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 200

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              |||| |   | ||   || ||||| | |  ||||| || |     || ||| |
Db        201 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 253


RESULT 14
US-09-252-991A-10818/c
; Sequence 10818, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10818
;   LENGTH: 1077
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10818

  Query Match             6.5%;  Score 29.8;  DB 4;  Length 1077;
  Best Local Similarity   54.0%;  Pred. No. 2.8;
  Matches   61;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
              | | || | |  |  || | |  | ||| |  |    ||||  ||||||  | |     |
Db        530 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 471

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              |||| |   | ||   || ||||| | |  ||||| || |     || ||| |
```

```
Db        470 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 418


RESULT 15
US-09-252-991A-10865
; Sequence 10865, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10865
;   LENGTH: 1521
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10865

  Query Match             6.5%;  Score 29.8;  DB 4;  Length 1521;
  Best Local Similarity   54.0%;  Pred. No. 3.3;
  Matches   61;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
              | | || | |  |  || | |  | ||| |  |    ||||  ||||||  | |     |
Db        170 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 229

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              |||| |   | ||   || ||||| | |  ||||| || |     || ||| |
Db        230 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 282


Search completed: March 15, 2004, 10:15:05
Job time : 74.8 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 07:24:24 ; Search time 340.6 Seconds
(without alignments)
4928.974 Million cell updates/sec

Title: US-09-543-407-27
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

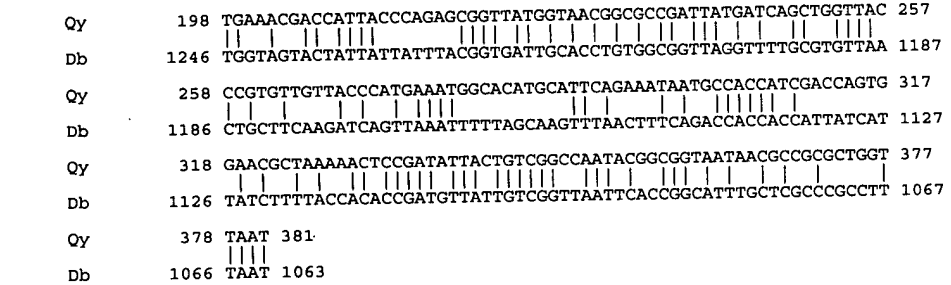
Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	36.8	8.1	5838	12	US-10-282-122A-26763	Sequence 26763, A
	2	34.6	7.6	1355	15	US-10-045-674-593	Sequence 593, App
c	3	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c	4	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c	5	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c	6	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c	7	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
c	8	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
	9	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	10	33	7.2	775	12	US-10-424-599-111819	Sequence 111819,
	11	33	7.2	1274	12	US-10-424-599-111822	Sequence 111822,
c	12	33	7.2	3390	12	US-10-282-122A-27533	Sequence 27533, A
	13	32.8	7.2	962	12	US-10-424-599-62799	Sequence 62799, A
c	14	32.4	7.1	1287	12	US-10-282-122A-7126	Sequence 7126, Ap
c	15	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
	16	32	7.0	1032	15	US-10-369-493-44704	Sequence 44704, A
	17	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
	18	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
	19	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c	20	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
	21	31.8	7.0	1668	15	US-10-355-956-3	Sequence 3, Appli
	22	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
	23	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
	24	31.6	6.9	1518	15	US-10-108-260A-2091	Sequence 2091, Ap
c	25	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
	26	31.4	6.9	1068	12	US-10-424-599-88687	Sequence 88687, A
c	27	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
c	28	31.2	6.8	960	12	US-10-282-122A-35869	Sequence 35869, A
	29	31.2	6.8	1416	15	US-10-388-934-99	Sequence 99, Appl
c	30	31	6.8	2142	14	US-10-128-714-2084	Sequence 2084, Ap
c	31	31	6.8	2169	14	US-10-128-714-7084	Sequence 7084, Ap
c	32	31	6.8	2221	14	US-10-128-714-1084	Sequence 1084, Ap
c	33	31	6.8	2345	14	US-10-128-714-6084	Sequence 6084, Ap
c	34	31	6.8	4221	14	US-10-128-714-84	Sequence 84, Appl
c	35	31	6.8	4345	14	US-10-128-714-5084	Sequence 5084, Ap
	36	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
c	37	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
	38	30.6	6.7	1011	12	US-10-282-122A-36185	Sequence 36185, A
	39	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
c	40	30.6	6.7	75216	14	US-10-080-170-646	Sequence 646, App
c	41	30.4	6.7	992	12	US-10-425-114-31747	Sequence 31747, A
c	42	30.4	6.7	3786	14	US-10-245-802-7	Sequence 7, Appli
	43	30.2	6.6	229	9	US-09-864-761-29677	Sequence 29677, A
	44	30.2	6.6	379	12	US-10-424-599-102524	Sequence 102524,
c	45	30.2	6.6	525	14	US-10-029-386-13465	Sequence 13465, A

ALIGNMENTS

RESULT 1
US-10-282-122A-26763/c
; Sequence 26763, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

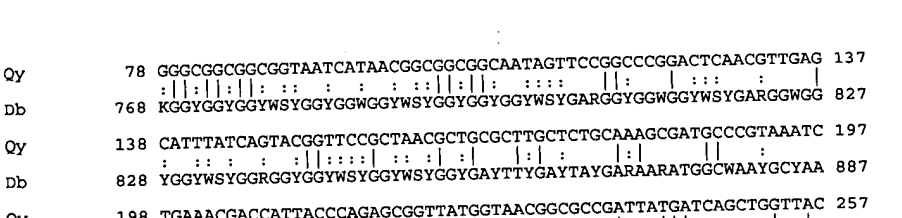
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26763
; LENGTH: 5838
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-26763

Query Match 8.1%; Score 36.8; DB 12; Length 5838;
Best Local Similarity 50.0%; Pred. No. 0.16;
Matches 92; Conservative 0; Mismatches 92; Indels 0; Gaps 0;



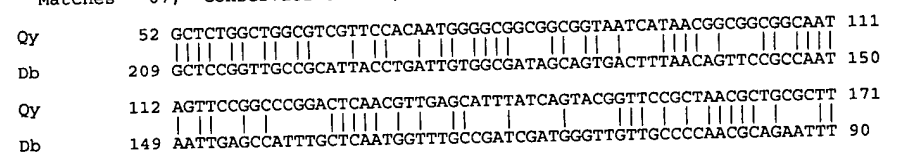
RESULT 2
US-10-045-674-593
; Sequence 593, Application US/10045674
; Publication No. US20030232333A1
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
US-10-045-674-593

Query Match 7.6%; Score 34.6; DB 15; Length 1355;
Best Local Similarity 24.1%; Pred. No. 0.45;
Matches 73; Conservative 83; Mismatches 147; Indels 0; Gaps 0;



RESULT 3
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.72;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;



RESULT 4
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay

; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.72;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db 209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy 112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db 149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy 172 GC 173
Db 89 GC 88

RESULT 5
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.72;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db 209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy 112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db 149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy 172 GC 173
Db 89 GC 88

RESULT 6
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-010-390-6

Query Match 7.5%; Score 34; DB 14; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.72;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db 209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy 112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db 149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy 172 GC 173
Db 89 GC 88

RESULT 7
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (EBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-441-736-5

Query Match 7.5%; Score 34; DB 15; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.72;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db 209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy 112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db 149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

```
Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 8
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1662)
US-10-156-761-3483

Query Match 7.3%; Score 33.2; DB 14; Length 1662;
Best Local Similarity 52.1%; Pred. No. 1.5;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               | || | ||  | ||||| | || |  | | || || || | ||| | ||||||   |
Db         590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531

Qy         110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
                 |  ||| ||  | | |   ||   ||  | |      |  || |||   | | |||
Db         530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

Qy         170 TTGCTCTGCAAAGCGATGCCCG 191
               |||  || |  |||  | | |
Db         470 TTGAACTTCTTGGCGGCGACGG 449
```

RESULT 9
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 7.3%; Score 33.2; DB 14; Length 9025608;
Best Local Similarity 52.1%; Pred. No. 1e+02;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               | || | ||  | ||||| | || |  | | || || || | ||| | ||||||   |
Db    4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

Qy         110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
                 |  ||| ||  | | |   ||   ||  | |      |  || |||   | | |||
Db    4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

Qy         170 TTGCTCTGCAAAGCGATGCCCG 191
               |||  || |  |||  | | |
Db    4331901 TTGAACTTCTTGGCGGCGACGG 4331922
```

RESULT 10
US-10-424-599-111819
; Sequence 111819, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 111819
; LENGTH: 775
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(775)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_71982C.1
US-10-424-599-111819

Query Match 7.2%; Score 33; DB 12; Length 775;
Best Local Similarity 44.3%; Pred. No. 1.2;
Matches 135; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

```
Qy          99 CGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGC 158
               || || |||| ||      || | |      | || |      |  |      | |||||
Db          94 CGCCGCCGGCTATCCCAACGTCTCTCCTCCTCCGCTCGCTGATCCTGTCGGTGCTTCCGA 153

Qy         159 TAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAG 218
               |  ||||||   |||   | |   || ||       || ||||||    |
Db         154 TCTCGCTGCTGCTGCCTCGGACGACGTTGTTGCGGAACCTGAAATTCACGCGCTGATTTC 213

Qy         219 CGGTTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAAT 278
               |  | |  |||     |||| | ||||  |  |   | ||    |  ||||||  |
Db         214 AGACGAAGAGCACGATCACGATCACGATCACGAGCATGAGCATGAGCATGACCATGACCT 273

Qy         279 GGCACATGCATTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATAT 338
                 | || |   |   || |     |  | || ||    || | |  || |||   |
```

```
Db        274 CGAGCACGATCACGAGCATGAGCATGACCACCACGAGGACCGCGATGATCACGCCGCCGT 333

Qy        339 TACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              | | | | | | | | || | | | |||| ||| | | | |||
Db        334 GGCAGCTGTCTCCTTGGAAGATCTCAAACTCAAGATCATTAAGCAGGTGGAGTATTATTT 393

Qy        399 CAGCG 403
              |||||
Db        394 CAGCG 398


RESULT 11
US-10-424-599-111822
; Sequence 111822, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 111822
;   LENGTH: 1274
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_71985C.1
US-10-424-599-111822

  Query Match             7.2%;  Score 33;  DB 12;  Length 1274;
  Best Local Similarity   44.3%;  Pred. No. 1.6;
  Matches  135;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy         99 CGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGC 158
              || || |||| ||     || | |     | || |     | |     | |||||
Db        133 CGCCGCCGGCTATCCCAACGTCTCTCCTCCTCCGCTCGCTGATCCTGTCGGTGCTTCCGA 192

Qy        159 TAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAG 218
              | ||||||   |||   | |   || ||       || ||||||    |
Db        193 TCTCGCTGCTGCTGCCTCGGACGACGTTGTTGCGGAACCTGAAATTCACGCGCTGATTTC 252

Qy        219 CGGTTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAAT 278
              |   | |   |||    |||| | |||||   |  |   | ||    |  ||||||  |
Db        253 AGACGAAGAGCACGATCACGATCACGATCACGAGCATGAGCATGAGCATGACCATGACCT 312

Qy        279 GGCACATGCATTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATAT 338
              |  || |    |    || |      |  | || ||     || | |  || |||   |
Db        313 CGAGCACGATCACGAGCATGAGCATGACCACCACGAGGACCGCGATGATCACGCCGCCGT 372

Qy        339 TACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              | | | | | | | | || | | | |||| ||| | | | |||
Db        373 GGCAGCTGTCTCCTTGGAAGATCTCAAACTCAAGATCATTAAGCAGGTGGAGTATTATTT 432

Qy        399 CAGCG 403
              |||||
Db        433 CAGCG 437


RESULT 12
US-10-282-122A-27533/c
; Sequence 27533, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27533
;   LENGTH: 3390
;   TYPE: DNA
;   ORGANISM: Mycobacterium leprae
US-10-282-122A-27533

  Query Match             7.2%;  Score 33;  DB 12;  Length 3390;
  Best Local Similarity   55.8%;  Pred. No. 2.6;
  Matches   63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              |||| |  || || | | |  ||||||||| |   |  | |||   ||   ||    ||
Db        869 ACGCCGACCTCGCGCAGGATCGCGATGCCCAGATCCCGCATACGTTGATATTCCCAGTCG 810

Qy        221 GTTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 273
              || |  || | ||||      || | | | |   | |||  ||||    ||||||
Db        809 GTCAGAGTCATGGCGGGTGCTACGGTGACCGAATCACCAGTGTGCACACCCAT 757


RESULT 13
US-10-424-599-62799
; Sequence 62799, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 62799
;   LENGTH: 962
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
```

```
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_27721C.1
US-10-424-599-62799

  Query Match             7.2%;  Score 32.8;  DB 12;  Length 962;
  Best Local Similarity   48.9%;  Pred. No. 1.6;
  Matches   88;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        249 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCCACCAT 308
              ||  |   |||| |   | |   | ||| |||  |  |  | ||   | ||  | |||| |
Db        782 GCGGTGAACCCTTTCAGCTCTTCCTGACATGTAATTTAAACTCGATATTATAGACACAAG 841

Qy        309 CGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGC 368
              |    || |||  |    |||| ||  || |||| |   |  | ||| |  | | |
Db        842 TTATATATGCAACTATTGCCACTCAGAATTTTCTGTGGCATATAAAGGTGTGATAATTGT 901

Qy        369 CGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGG 428
                | || ||| ||    | | | ||| || |  || || | |   ||  ||   |||
Db        902 TTTCCATGTAAATTAGTTATCTTATAATTGCAACCAAAGGGGGGGGGGGGGGGGGGTGG 961


RESULT 14
US-10-282-122A-7126/c
; Sequence 7126, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7126
;   LENGTH: 1287
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-282-122A-7126

  Query Match             7.1%;  Score 32.4;  DB 12;  Length 1287;
  Best Local Similarity   54.1%;  Pred. No. 2.5;
  Matches   66;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        190 CGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTATGATCAG 249
              ||  || | | |   | ||| | |||  || | |  |||| |  ||  | |   | ||||
Db        447 CGGTAAACCGTAGAAAGCATGAGCCACTGCCGCTCCGGTATCACCGGAGGTCGCGGTCAG 388

Qy        250 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCCACCATC 309
                | | | |  || | ||||  | ||||  ||    ||  |  |||| || ||| |
Db        387 AATGGTCACTGGCTTATCACCCGCAATATGGGTCAGCATTTGTGCCATAAAGCGACCGCC 328

Qy        310 GA 311
              ||
Db        327 GA 326


RESULT 15
US-10-369-493-33133/c
; Sequence 33133, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33133
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Xylella fastidiosa
US-10-369-493-33133

  Query Match             7.1%;  Score 32.2;  DB 15;  Length 1083;
  Best Local Similarity   59.1%;  Pred. No. 2.7;
  Matches   55;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTG 421
              ||  | ||| |||||    |  | |||||     ||  || | ||| | ||| ||||
Db        488 ATCTCACCGAGCTGGGATAAAAAACCGCGCACAATCACACCATAACGCTGCTTCAGCCAC 429

Qy        422 GTTTTGGCAACAACGCCACGGCTAACCAGTATT 454
              |||| ||||| || |||    | ||  | |||
Db        428 TTTTTAGCAACCACCCCAGCAGTCACACGCATT 396


Search completed: March 15, 2004, 14:16:36
Job time : 352.6 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:52:04 ; Search time 2500.2 Seconds
(without alignments)
6417.435 Million cell updates/sec

Title: US-09-543-407-27
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-27	Sequence 27, Appl
2	401.6	88.1	456	6	US-08-233-642A-56	Sequence 56, Appl
3	401.6	88.1	456	23	US-09-543-407-1	Sequence 1, Appli
4	376	82.5	456	23	US-09-543-407-21	Sequence 21, Appl
5	360	78.9	456	23	US-09-543-407-19	Sequence 19, Appl
6	352	77.2	456	23	US-09-543-407-17	Sequence 17, Appl
7	350.4	76.8	456	23	US-09-543-407-11	Sequence 11, Appl
8	350.4	76.8	456	23	US-09-543-407-13	Sequence 13, Appl
9	350.4	76.8	456	23	US-09-543-407-23	Sequence 23, Appl
10	350.4	76.8	456	23	US-09-543-407-25	Sequence 25, Appl
11	348.8	76.5	456	23	US-09-543-407-15	Sequence 15, Appl
12	344	75.4	456	23	US-09-543-407-29	Sequence 29, Appl
13	277.4	60.8	361	6	US-08-233-642A-54	Sequence 54, Appl
14	228.8	50.2	456	23	US-09-543-407-3	Sequence 3, Appli
15	228.8	50.2	477	17	US-09-252-691-172	Sequence 172, App
16	228.8	50.2	477	17	US-09-252-691C-172	Sequence 172, App
17	228.8	50.2	477	51	US-10-417-886-172	Sequence 172, App
18	228.8	50.2	675	13	US-08-978-878-3	Sequence 3, Appli
19	228.8	50.2	675	31	US-09-741-873B-3	Sequence 3, Appli
20	226.4	49.6	453	92	US-60-352-946-1	Sequence 1, Appli
21	226.4	49.6	453	101	US-60-444-371-1	Sequence 1, Appli
22	184.8	40.5	396	13	US-08-978-878-1	Sequence 1, Appli
23	160	35.1	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
c 28	49.8	10.9	78	23	US-09-543-407-52	Sequence 52, Appl
29	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
30	48	10.5	78	23	US-09-543-407-53	Sequence 53, Appl
c 31	46.6	10.2	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
32	41.8	9.2	32396	36	US-09-897-516-112	Sequence 112, App
c 33	41.8	9.2	32396	36	US-09-897-516-273	Sequence 273, App
34	41.8	9.2	32396	36	US-09-897-516-484	Sequence 484, App
c 35	41.8	9.2	32396	36	US-09-897-516-485	Sequence 485, App
36	41.8	9.2	32396	36	US-09-897-516-3259	Sequence 3259, Ap
37	41.8	9.2	32396	36	US-09-897-516-3260	Sequence 3260, Ap
38	41.8	9.2	32396	36	US-09-897-516-3261	Sequence 3261, Ap
39	41.8	9.2	32396	36	US-09-897-516-3262	Sequence 3262, Ap
40	41.8	9.2	32396	36	US-09-897-516-3263	Sequence 3263, Ap
41	41.8	9.2	32396	36	US-09-897-516-3264	Sequence 3264, Ap
42	41.8	9.2	32396	36	US-09-897-516-3265	Sequence 3265, Ap
43	41.8	9.2	32396	36	US-09-897-516-3266	Sequence 3266, Ap
44	41.8	9.2	32396	36	US-09-897-516-3267	Sequence 3267, Ap
45	41.8	9.2	32396	36	US-09-897-516-3268	Sequence 3268, Ap

ALIGNMENTS

```
RESULT 1
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-136;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;     APPLICANT:  Kay, William W.
;     APPLICANT:  Collinson, S. Karen
;     APPLICANT:  Clouthier, Sharon C.
;     APPLICANT:  Doran, James L.
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;     TITLE OF INVENTION:  BASED VACCINES
;     NUMBER OF SEQUENCES:  58
;     NUMBER OF SEQUENCES:  -
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Seed and Berry
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  U.S.A.
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/233,642A
;       FILING DATE:  26-APR-1994
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  King, Joshua
```

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.1%;  Score 401.6;  DB 6;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 2.1e-118;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                |  |||  |   |  |  ||  |   |    |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.1%;  Score 401.6;  DB 23;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 2.1e-118;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                |  |||  |   |  |  ||  |   |    |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             82.5%;  Score 376;  DB 23;  Length 456;
  Best Local Similarity   89.0%;  Pred. No. 3.9e-110;
  Matches  406;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy       241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
Db       241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
Qy       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

Query Match 78.9%; Score 360; DB 23; Length 456;
Best Local Similarity 86.8%; Pred. No. 5.7e-105;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
Qy       241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
Db       241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.2e-102;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
Qy       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy       241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
Db       241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 7.2e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              | |||  |      | |  ||  |   |     ||| ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 7.2e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              | |||  |      | |  ||  |   |     ||| ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 7.2e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              |  |||  |       |  |  ||  |    |     |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||     ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||  |   ||  |  |||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 7.2e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||  |  ||  | ||  |||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              | |||  |     |  | ||  |    |     |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 2.4e-101;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||  ||  | | | |  | |     |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                |  |||  |      |  |  ||  |    |      |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 75.4%; Score 344; DB 23; Length 456;
Best Local Similarity 84.6%; Pred. No. 8.4e-100;
Matches 386; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                |  |||  |      |  |  ||  |    |      |||  ||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |    |                |  | |  ||   ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 60.8%; Score 277.4; DB 6; Length 361;
Best Local Similarity 89.3%; Pred. No. 2.5e-78;
Matches 299; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTAT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCC 303
              |  |||  |      |  |  ||   |    |      |||  ||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. Coli
US-09-543-407-3

Query Match 50.2%; Score 228.8; DB 23; Length 456;
Best Local Similarity 68.9%; Pred. No. 1.4e-62;
Matches 314; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || |||||  || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||  ||   ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              || ||||                 | | ||  |   || ||  ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

RESULT 15
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-172

Query Match 50.2%; Score 228.8; DB 17; Length 477;
Best Local Similarity 68.9%; Pred. No. 1.4e-62;
Matches 314; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||  |||||||||    | ||||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || ||||||||  |||||||||      |  |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  ||||||||||  ||||||||    || ||||  | |||||||||  ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | |||||| ||||||||||||   |||  ||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
              || |||      |       | || | ||    ||  | | |  ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||| |||||   ||| |||| | |||  |  ||   || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||||  |||  | |||||  || ||  || |  || |  |||| ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  ||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477
```

Search completed: March 15, 2004, 13:42:50
Job time : 2501.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 04:16:29 ; Search time 53.4 Seconds
(without alignments)
2945.701 Million cell updates/sec

Title: US-09-543-407-27
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 628836 seqs, 172478556 residues

Total number of hits satisfying chosen parameters: 1257672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	228.8	50.2	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	184.8	40.5	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	34.6	7.6	1355	6	US-10-045-674A-593	Sequence 593, App
c	4	31.8	7.0	3060	7	US-60-544-461-28	Sequence 28, Appl
c	5	31.8	7.0	3060	7	US-60-540-661-28	Sequence 28, Appl
c	6	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
c	7	29.8	6.5	1280	6	US-10-644-123-2	Sequence 2, Appli
c	8	29.8	6.5	1397	6	US-10-644-123-1	Sequence 1, Appli
	9	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
	10	29.4	6.4	350	6	US-10-767-701-19256	Sequence 19256, A
c	11	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	12	29.2	6.4	619	6	US-10-767-701-1967	Sequence 1967, Ap
	13	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
	14	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	15	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
	16	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	17	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	18	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
	19	28.2	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	20	28.2	6.2	519599	6	US-10-765-790-73	Sequence 73, Appl
	21	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
	22	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
	23	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
	24	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
c	25	27.6	6.1	2443	5	US-09-973-180A-1	Sequence 1, Appli
	26	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	27	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	28	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
	29	27.4	6.0	384	6	US-10-451-467A-677	Sequence 677, App
c	30	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
	31	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
	32	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
c	33	27.2	6.0	861	6	US-10-767-701-3616	Sequence 3616, Ap
c	34	27.2	6.0	2880	6	US-10-100-683-5057	Sequence 5057, Ap
c	35	27.2	6.0	2881	6	US-10-100-683-2682	Sequence 2682, Ap
c	36	27.2	6.0	7380	6	US-10-417-375A-23	Sequence 23, Appl
c	37	27.2	6.0	7400	6	US-10-417-375A-19	Sequence 19, Appl
c	38	27.2	6.0	7460	6	US-10-417-375A-25	Sequence 25, Appl
c	39	27.2	6.0	113486	6	US-10-417-375A-18	Sequence 18, Appl
c	40	27.2	6.0	510510	7	US-60-548-091-5616	Sequence 5616, Ap
	41	27	5.9	594	6	US-10-767-701-15041	Sequence 15041, A
	42	27	5.9	600	7	US-60-545-213-3064	Sequence 3064, Ap
	43	27	5.9	600	7	US-60-545-213-7336	Sequence 7336, Ap
	44	27	5.9	725	6	US-10-767-701-8779	Sequence 8779, Ap
c	45	27	5.9	745	6	US-10-779-543-3681	Sequence 3681, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 50.2%; Score 228.8; DB 5; Length 675;
Best Local Similarity 68.9%; Pred. No. 1.1e-68;
Matches 314; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873C
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873C-1

  Query Match             40.5%;  Score 184.8;  DB 5;  Length 396;
  Best Local Similarity   66.7%;  Pred. No. 1e-53;
  Matches  264;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396


RESULT 3
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674A
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 06/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674A-593

  Query Match             7.6%;  Score 34.6;  DB 6;  Length 1355;
  Best Local Similarity   24.1%;  Pred. No. 0.051;
  Matches   73;  Conservative   83;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137

Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197

Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTATGATCAGCTGGTTAC 257

Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        258 CCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCCACCATCGACCAGTG 317

Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYATYGGYGAYGTYWS 1007

Qy        318 GAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGT 377

Db       1008 YGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKAATTCYCARATGGC 1067
```

```
Qy        378 TAA 380
              : |
Db       1068 YCA 1070


RESULT 4
US-60-544-461-28/c
; Sequence 28, Application US/60544461
; GENERAL INFORMATION:
;  APPLICANT: Brown, Kimberly
;  APPLICANT:  Harris, Paul
;  APPLICANT:  Zaretsky, Elizabeth J.
;  APPLICANT:  Rey, Edward
;  APPLICANT:  Vlasenko, Elena
;  APPLICANT:  McFarland, Keith C.
;  TITLE OF INVENTION: Polypeptides Having Cellulolytic Enhancing Activity And Nucleic A
;  TITLE OF INVENTION:  Encoding Same
;  FILE REFERENCE: 10587.000-US
;  CURRENT APPLICATION NUMBER: US/60/544,461
;  CURRENT FILING DATE:  2004-02-12
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
;   LENGTH: 3060
;   TYPE: DNA
;   ORGANISM: Aspergillus fumigatus
US-60-544-461-28

  Query Match             7.0%;  Score 31.8;  DB 7;  Length 3060;
  Best Local Similarity   48.6%;  Pred. No. 0.68;
  Matches  120;  Conservative    0;  Mismatches  122;  Indels    5;  Gaps    1;

Qy         97 AACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCC 156
              |||| | ||  | || ||||      ||   ||  ||| | |    | |     |||||
Db       1080 AACGCCAGCTACGATGGTTCGTCAGCGATTTCTTCGTCGCGAGGTCAAGTCTACGGAAAA 1021

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCG-----ATGCCCGTAAATCTGAAACGACCATTA 211
               || ||   || |    |||||||||  |      || || |    |     || | | ||
Db       1020 TCTTACCGCGCACAGCATCTGCAAAGGGCCTGCATTCCAGGTCAGTCAAATCAATCAAGG 961

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCC 271
              ||  |   |   ||  |||  | |     | |  || || ||     || |||| | ||
Db        960 TCCTCATTTGCAGTGTCAACTACTCACCAAAGGTACAACTCGTGCATGGTCTTGTCATCC 901

Qy        272 ATGAAATGGCACATGCATTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              | |      ||  ||    || |  ||  ||  |||||   || | |   ||| ||     |
Db        900 ACGTTGGAGCTGATCGTCTCCGTGATGTTGTAACCATATCCCTGGGCCTCGCCAACCTGT 841

Qy        332 CCGATAT 338
              | || ||
Db        840 CGGAAAT 834


RESULT 5
US-60-540-661-28/c
; Sequence 28, Application US/60540661
; GENERAL INFORMATION:
;  APPLICANT: Brown, Kimberly
;  APPLICANT:  Harris, Paul
;  APPLICANT:  Zaretsky, Elizabeth J.
;  APPLICANT:  Rey, Edward
;  APPLICANT:  Vlasenko, Elena
;  APPLICANT:  McFarland, Keith C.
;  TITLE OF INVENTION: Polypeptides Having Cellulolytic Enhancing Activity And Nucleic A
;  TITLE OF INVENTION:  Encoding Same
;  FILE REFERENCE: 10587.000-US
;  CURRENT APPLICATION NUMBER: US/60/540,661
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
;   LENGTH: 3060
;   TYPE: DNA
;   ORGANISM: Aspergillus fumigatus
US-60-540-661-28

  Query Match             7.0%;  Score 31.8;  DB 7;  Length 3060;
  Best Local Similarity   48.6%;  Pred. No. 0.68;
  Matches  120;  Conservative    0;  Mismatches  122;  Indels    5;  Gaps    1;

Qy         97 AACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCC 156
              |||| | ||  | || ||||      ||   ||  ||| | |    | |     |||||
Db       1080 AACGCCAGCTACGATGGTTCGTCAGCGATTTCTTCGTCGCGAGGTCAAGTCTACGGAAAA 1021

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCG-----ATGCCCGTAAATCTGAAACGACCATTA 211
               || ||   || |    |||||||||  |      || || |    |     || | | ||
Db       1020 TCTTACCGCGCACAGCATCTGCAAAGGGCCTGCATTCCAGGTCAGTCAAATCAATCAAGG 961

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCC 271
              ||  |   |   ||  |||  | |     | |  || || ||     || |||| | ||
Db        960 TCCTCATTTGCAGTGTCAACTACTCACCAAAGGTACAACTCGTGCATGGTCTTGTCATCC 901

Qy        272 ATGAAATGGCACATGCATTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              | |      ||  ||    || |  ||  ||  |||||   || | |   ||| ||     |
Db        900 ACGTTGGAGCTGATCGTCTCCGTGATGTTGTAACCATATCCCTGGGCCTCGCCAACCTGT 841

Qy        332 CCGATAT 338
              | || ||
Db        840 CGGAAAT 834


RESULT 6
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
;  APPLICANT: Bayer Pharmaceuticals Corporation
;  APPLICANT:  Eveleigh, Deepa
;  APPLICANT:  Bigwood, Douglas
;  TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
;  FILE REFERENCE: 5152
;  CURRENT APPLICATION NUMBER: US/10/788,792
;  CURRENT FILING DATE:  2004-02-27
;  PRIOR APPLICATION NUMBER: US 60/450,655
;  PRIOR FILING DATE: 2003-02-28
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
;   LENGTH: 1575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-788-792-27

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 1575;
  Best Local Similarity   55.7%;  Pred. No. 1.1;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
              |||   | | | | || ||   | | || |  || | |||    || |||| ||||||||
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
              | | |   | | |||  |  |   |   |||  |  ||||||||
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583


RESULT 7
US-10-644-123-2/c
; Sequence 2, Application US/10644123
; GENERAL INFORMATION:
;  APPLICANT: Richards, Nigel Gordon John
;  APPLICANT:  Chang, Christopher Harry
;  APPLICANT:  Peck, Ammon B.
;  TITLE OF INVENTION: Polunucleotides Encoding Oxalate Decarboxylase from Aspergillus
;  TITLE OF INVENTION:  Niger and Methods of Use
;  FILE REFERENCE: UF-314XC1
```

```
;  CURRENT APPLICATION NUMBER: US/10/644,123
;  CURRENT FILING DATE:  2003-08-20
;  PRIOR APPLICATION NUMBER: US 60/404,892
;  PRIOR FILING DATE: 2002-08-20
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 1280
;   TYPE: DNA
;   ORGANISM: Aspergillus niger
US-10-644-123-2

  Query Match             6.5%;  Score 29.8;  DB 6;  Length 1280;
  Best Local Similarity   55.2%;  Pred. No. 2.2;
  Matches   58;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        244 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCC 303
              || |  || || | |  ||  ||| | |||||||| | ||||     || || ||||||
Db       1093 GACCTCCTCGTCATCACTGAGGTTCTCAATGAAATGACCCATGTTGCGAGGAACAATGCC 1034

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGC 348
                ||||  ||  |     |   ||    |    |||||  || ||
Db       1033 CACATCTCCCGCTACGTAGTCGAATGTACGAGCATTACCTTCAGC 989


RESULT 8
US-10-644-123-1/c
; Sequence 1, Application US/10644123
; GENERAL INFORMATION:
;  APPLICANT: Richards, Nigel Gordon John
;  APPLICANT:  Chang, Christopher Harry
;  APPLICANT:  Peck, Ammon B.
;  TITLE OF INVENTION: Polunucleotides Encoding Oxalate Decarboxylase from Aspergillus
;  TITLE OF INVENTION:  Niger and Methods of Use
;  FILE REFERENCE: UF-314XC1
;  CURRENT APPLICATION NUMBER: US/10/644,123
;  CURRENT FILING DATE:  2003-08-20
;  PRIOR APPLICATION NUMBER: US 60/404,892
;  PRIOR FILING DATE: 2002-08-20
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1397
;   TYPE: DNA
;   ORGANISM: Aspergillus niger
US-10-644-123-1

  Query Match             6.5%;  Score 29.8;  DB 6;  Length 1397;
  Best Local Similarity   55.2%;  Pred. No. 2.3;
  Matches   58;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        244 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCC 303
              || |  || || | |  ||  ||| | |||||||| | ||||     || || ||||||
Db       1210 GACCTCCTCGTCATCACTGAGGTTCTCAATGAAATGACCCATGTTGCGAGGAACAATGCC 1151

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGC 348
                ||||  ||  |     |   ||    |    |||||  || ||
Db       1150 CACATCTCCCGCTACGTAGTCGAATGTACGAGCATTACCTTCAGC 1106


RESULT 9
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 1.6;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||     |       ||||| | |     | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
                 |||||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 10
US-10-767-701-19256
; Sequence 19256, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 19256
;   LENGTH: 350
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(350)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3480-035-P1-K1-B6
US-10-767-701-19256

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 350;
  Best Local Similarity   49.7%;  Pred. No. 1.7;
  Matches   75;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||||  || ||| | |  | || ||||| ||||  |       |  |||| |    | ||
Db         90 CAAAGAGAAGCCAGCATCTGTGCAACGAGCATTGTCTGTTCTTGAGATGGAAGTTGAGCA 149

Qy        238 GATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAAT 297
              || | ||   |  | |  |   ||  | |  |   ||   ||    ||  || | |
Db        150 AATAAAGAAAGGAAGTTATGACCACTTCATGCAAAAAGAAATCCATGAACAGCCACATTC 209

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAA 328
              | ||  | || | |  || || |  || |||
Db        210 ATTGAAAACAACAATGAGGGGTAGACTGAAA 240


RESULT 11
US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
```

```
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-117

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 249;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   |||||||||      || || ||| |    | |  | |   || | | |||
Db        180 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 121

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    | | ||  || || || ||| | | |||    |||| | | |    || |
Db        120 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 61

Qy        415 CA 416
              |
Db         60 GA 59


RESULT 12
US-10-767-701-1967
; Sequence 1967, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 1967
;   LENGTH: 619
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS33453_1
US-10-767-701-1967

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 619;
  Best Local Similarity   47.0%;  Pred. No. 2.5;
  Matches  126;  Conservative    0;  Mismatches  138;  Indels    4;  Gaps    1;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGG----TAATCATAACGG 101
              | | |||||| | |    |       |||   | ||||||||||    | ||||   |
Db        249 GCCGGTGCTCCGCCCCCCCCGTCCTCACCGGGCGGCGGCGGCGGCGGCTCATCAGCGCCA 308

Qy        102 CGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAA 161
              | |   |     |  ||||  |  ||| |      ||   | |  | |      |   | |
Db        309 CCGTACCGCCGTTGACGGCGTGCTCTCCAACGACGGCGGCTGTTGGAAGAAGCCGCGTGA 368

Qy        162 CGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGG 221
              || ||||||   || || ||||||     ||  ||||     | | | ||| ||| | ||
Db        369 CGTTGCGCTCCGTCGGCGAAGCGAGAAGCGAGAATCAAGTCCCATCGTTAGCCACATCAA 428

Qy        222 TTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGC 281
              | || |    | |||   | ||    |  ||   |  ||| | |   ||    |   |
Db        429 TCATCGATCGGTCGCGATTGATAGGGAAGTGTACAGTCGTATAGCGTCCTCCCAGCAGAA 488

Qy        282 ACATGCATTCAGAAATAATGCCACCATC 309
                |       |||  || ||  || | ||
Db        489 TCCCTGCGCCAGTCATTATTTCAGCTTC 516


RESULT 13
US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 757
;   TYPE: DNA
;   ORGANISM: E. Coli
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(757)
;   OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 757;
  Best Local Similarity   52.5%;  Pred. No. 2.8;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   |||||||||      || || ||| |    | |  | |   || | | |||
Db        264 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 323

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    | | ||  || || || ||| | | |||    |||| | | |    || |
Db        324 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 383

Qy        415 CA 416
              |
Db        384 GA 385


RESULT 14
PCT-US04-05654-1030
; Sequence 1030, Application PC/TUS0405654
; GENERAL INFORMATION:
;  APPLICANT: Sherman, Bradley K
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline  E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Dubell III, Arnold N
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Repetti, Peter
;  APPLICANT:  Century, Karen
;  APPLICANT:  Gutterson, Neal
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Kumimoto, Roderick W
;  APPLICANT:  Pilgrim, Marsha L
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 PCT
;  CURRENT APPLICATION NUMBER: PCT/US04/05654
```

```
;  CURRENT FILING DATE:  2004-03-03
;  PRIOR APPLICATION NUMBER: 10/374,780
;  PRIOR FILING DATE: 2003-02-25
;  PRIOR APPLICATION NUMBER: 10/675,852
;  PRIOR FILING DATE: 2003-09-30
;  NUMBER OF SEQ ID NOS: 2950
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1030
;   LENGTH: 1186
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
PCT-US04-05654-1030

  Query Match             6.3%;  Score 28.6;  DB 1;  Length 1186;
  Best Local Similarity   59.0%;  Pred. No. 5.5;
  Matches   49;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         44 CTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCG 103
              | || | |||||  || | || ||      |  | || ||||||||  |  |  ||||
Db        606 CAGGGACTGCTCCAGCGGCCGGCGGCGGCTACGGCGGTGGCGGCGGCGGTGGTGGTGGCG 665

Qy        104 GCGGCAATAGTTCCGGCCCGGAC 126
              || || | |  | ||||| || |
Db        666 GCTGCTACAACTGCGGCCAGGCC 688


RESULT 15
US-60-548-091-22688
; Sequence 22688, Application US/60548091
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STROKE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001506
;  CURRENT APPLICATION NUMBER: US/60/548,091
;  CURRENT FILING DATE:  2004-02-27
;  NUMBER OF SEQ ID NOS: 24433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22688
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-548-091-22688

  Query Match             6.2%;  Score 28.2;  DB 7;  Length 201;
  Best Local Similarity   62.7%;  Pred. No. 3.3;
  Matches   42;  Conservative    1;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        360 TAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGT 419
              |||||  | |   |  |||||   | |   |:|||| |  | ||| ||||  ||||||  a
Db         69 TAATATTGTCAGTCATCTTAATTTTAGCCATTYTGATACATGTGTAGTGGTATGTCAGGG 128

Qy        420 TGGTTTT 426
              |||||||
Db        129 TGGTTTT 135


Search completed: March 15, 2004, 13:47:28
Job time : 55.4 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 02:55:53 ; Search time 1963.4 Seconds
(without alignments)
6935.497 Million cell updates/sec

Title: US-09-543-407-27
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	229.4	50.3	680	12	BJ618688	BJ618688 BJ618688
	2	37.8	8.3	725	14	CD493501	CD493501 CDA05-E04
c	3	37.6	8.2	699	28	CC098707	CC098707 CSU-K34.1
c	4	37.4	8.2	484	9	AV430994	AV430994 AV430994
c	5	37.4	8.2	610	14	CK130514	CK130514 AT09365.3
	6	36.2	7.9	768	12	BI667482	BI667482 603292284
	7	36	7.9	479	9	AA263424	AA263424 LD07239.5
	8	36	7.9	545	12	BI371815	BI371815 RE58816.5
	9	36	7.9	627	12	BI162493	BI162493 RE01790.5
c	10	35.8	7.9	543	12	BG632414	BG632414 GH08173.3
c	11	35.8	7.9	578	10	AW944654	AW944654 SD10674.3
c	12	35.8	7.9	659	14	CA806764	CA806764 ESG014d.E
c	13	35.8	7.9	678	14	CA804002	CA804002 ESG011c.
c	14	35.8	7.9	723	14	CA807388	CA807388 ESG017c.E
c	15	35.8	7.9	800	12	BG633139	BG633139 GH20822.3
	16	35.4	7.8	359	10	BE942582	BE942582 EST422161
	17	35.4	7.8	469	10	BE942583	BE942583 EST422162
	18	35.4	7.8	736	12	BI310708	BI310708 EST531245
	19	35.4	7.8	814	12	BG586907	BG586907 EST488676
	20	35.4	7.8	942	29	CNS00I0L	AL074582 Drosophil
	21	35.2	7.7	393	12	BI587644	BI587644 RH28379.5
	22	35.2	7.7	539	14	CA506913	CA506913 UI-R-FJ0-
	23	35.2	7.7	554	12	BI367839	BI367839 RE53687.5
	24	34.6	7.6	550	29	BX216179	BX216179 Danio rer
	25	34.6	7.6	804	29	BX235146	BX235146 Danio rer
	26	34.4	7.5	379	12	BI579559	BI579559 RE73612.5
	27	34.4	7.5	383	9	AA201611	AA201611 LD04667.5
	28	34.4	7.5	383	12	BI586927	BI586927 RH27154.5
	29	34.4	7.5	389	12	BI212591	BI212591 RE17708.5
	30	34.4	7.5	408	12	BI242349	BI242349 RE39652.5
	31	34.4	7.5	410	12	BI243539	BI243539 RE41205.5
	32	34.4	7.5	418	12	BI592772	BI592772 RH10552.5
	33	34.4	7.5	422	12	BI375702	BI375702 RE63566.5
	34	34.4	7.5	425	12	BI239173	BI239173 RE35589.5
	35	34.4	7.5	426	12	BI212443	BI212443 RE17536.5
	36	34.4	7.5	428	12	BI367567	BI367567 RE53360.5
	37	34.4	7.5	430	12	BI235073	BI235073 RE30876.5
	38	34.4	7.5	437	12	BI588763	BI588763 RH30178.5
	39	34.4	7.5	445	12	BI243839	BI243839 RE41624.5
	40	34.4	7.5	446	12	BI164458	BI164458 RE04136.5
	41	34.4	7.5	451	12	BI593541	BI593541 RH11637.5
	42	34.4	7.5	452	12	BI167024	BI167024 RE07294.5
	43	34.4	7.5	458	9	AA392334	AA392334 LD11173.5
	44	34.4	7.5	459	9	AI063597	AI063597 GH03503.5
	45	34.4	7.5	461	9	AA803227	AA803227 GM10427.5

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

```
            http://xenopus.nibb.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match             50.3%;  Score 229.4;  DB 12;  Length 680;
  Best Local Similarity   68.9%;  Pred. No. 3.1e-55;
  Matches  314;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||   | ||    | ||| |||||| |||||||||   || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||   ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
                || ||||                | | ||    |    || ||  ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACCGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138


RESULT 2
CD493501
LOCUS       CD493501                725 bp    mRNA    linear   EST 12-JUN-2003
DEFINITION  CDA05-E04.x1d-t SHGC-CDA Gasterosteus aculeatus cDNA clone
            CDA05-E04 5', mRNA sequence.
ACCESSION   CD493501
VERSION     CD493501.1  GI:31420532
KEYWORDS    EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
  ORGANISM  Gasterosteus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M.,
            Schmutz,J. and Myers,R.M.
  TITLE     Expressed sequence tags from Gasterosteus aculeatus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Kingsley, DM
            HHMI and Department of Developmental Biology
            Stanford University School of Medicine
            Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
            Tel: 650 725 5954
            Fax: 650 725 7739
            Email: kingsley@cmgm.stanford.edu
            Plate: 05
            High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Gasterosteus aculeatus"
                     /mol_type="mRNA"
                     /strain="Salinas river, CA"
                     /db_xref="taxon:69293"
                     /clone="CDA05-E04"
                     /sex="mixed male and female"
                     /tissue_type="heads and internal organs combined"
                     /dev_stage="adult"
                     /clone_lib="SHGC-CDA"
                     /note="Vector: lambda ZAP Express/pBK-CMV; Site_1: EcoR1
                     (5' adaptor); Site_2: XhoI (3' linker primer); The mixed
                     organ cDNA library was generated using the ZAP-cDNA method
                     by Stratagene. First strand cDNA synthesis was primed with
                     a a 50 bp linker primer containing an oligo dT sequence
                     preceeded by a synthetic XhoI site. 5 prime adaptors were
                     used containing an EcoR1 cohesive end. The finished cDNAs
                     were inserted in to the ZAP express vector
                     unidirectionally in the sense orientation with respect to
                     the lacZ promoter of pBK-CMV. An amplified library was
                     prepared from approximately 3 million primary clones in
                     the lambda ZAP Express vector. In vivo excision was then
                     used to generate individual pBK-CMV phagemid clones for
                     EST sequencing."
ORIGIN

  Query Match              8.3%;  Score 37.8;  DB 14;  Length 725;
  Best Local Similarity   52.9%;  Pred. No. 7.7;
  Matches   81;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        253 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAATGCCACCATCGAC 312
              || ||  | ||||  |  || ||  || |  | ||  ||||||  |  |    ||||| |
Db         47 GTAACTTGCGTTGACAGACAGGATTTGACCGAGGCTCTCAGAATTTTTCTGGCCATCCAG 106

Qy        313 CAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCG 372
              ||   ||| | |||| ||    |  | || | | |     ||| || | |  |
Db        107 GAGCCAAACTCCAAAATCTATTTTGACATTGGCTGTCTCCATCTTCTTAACAATGACCTG 166

Qy        373 CTGGTTAATCAGACCGCATCTGATTCCAGCGTA 405
                | |  | | |  || | ||||| |||| ||
Db        167 GATGCTGCTGAAAAGGCGTTTGATTGCAGCATA 199


RESULT 3
CC098707/c
LOCUS       CC098707                699 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K34.111B24.SP6 CSU-K34 Aedes aegypti genomic clone
            CSU-K34.111B24, genomic survey sequence.
ACCESSION   CC098707
VERSION     CC098707.1  GI:29964577
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K34.111B24.T7
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
```

```
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .699
                      /organism="Aedes aegypti"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7159"
                      /clone="CSU-K34.111B24"
                      /clone_lib="CSU-K34"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Source DNA: Aedes
                      aegypti; strain unknown [derived from freshly hatched
                      larvae at the Virus Research Centre, Poona, India.
                      Reference: SINGH, K. R. P., 1967 Cell cultures derived
                      from larvae of Aedes albopictus (Skuse) and Aedes aegypti
                      (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC
                      CCL-125"
ORIGIN

  Query Match             8.2%;  Score 37.6;  DB 28;  Length 699;
  Best Local Similarity   53.4%;  Pred. No. 8.7;
  Matches   79;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        132 GTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCG 191
              || | |  |  ||    | ||   ||||| ||||||||||||||| |   | |    ||
Db        162 GTAGTGTGTAAATGTCAAAGGCGACGCTAGCGCTGCGCTTGCTCGGTTGACCACCAACCA 103

Qy        192 TAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATTATGATCAGCT 251
              || ||  ||  ||| | |||   ||   |   |||| ||     | || | |||    |
Db        102 TATATTCGATTCGATCGTTAAATAGTTGGTCGATGGAAAATAATCAGAGTGTGACGTACG 43

Qy        252 GGTTACCCGTGTTGTTACCCATGAAATG 279
              | ||  | ||| |  | |    | ||||
Db         42 GTTTGTCTGTGCTTATTCTATCGCAATG 15


RESULT 4
AV430994/c
LOCUS       AV430994                484 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PL027e10_r 5', mRNA sequence.
ACCESSION   AV430994
VERSION     AV430994.1  GI:8586219
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
   PUBMED   10907854
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES              Location/Qualifiers
     source           1. .484
                      /organism="Porphyra yezoensis"
                      /mol_type="mRNA"
                      /strain="TU-1"
                      /db_xref="taxon:2788"
                      /clone="PL027e10_r"
                      /clone_lib="Porphyra yezoensis TU-1"
                      /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                      XhoI"
ORIGIN

  Query Match             8.2%;  Score 37.4;  DB 9;  Length 484;
  Best Local Similarity   57.1%;  Pred. No. 8.4;
  Matches   68;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |   | |  | | ||  ||| |||||   ||| ||||| ||||||  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
              | ||  ||   ||  ||||  ||| ||||||    |  ||   |||  |  || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323


RESULT 5
CK130514/c
LOCUS       CK130514                610 bp    mRNA    linear   EST 02-DEC-2003
DEFINITION  AT09365.3prime AT Drosophila melanogaster adult testes pOTB7
            Drosophila melanogaster cDNA clone AT09365 3 similar to CG4526:
            FBan0004526 GO:[transporter (GO:0005215)] located on: 3L
            73A3-73A4;: 08/14/2002, mRNA sequence.
ACCESSION   CK130514
VERSION     CK130514.1  GI:38622450
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E.,
            Berman,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V.,
            Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N.,
            Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C.,
            Lewis,S.E., Celniker,S. and Rubin,G.M.
  TITLE     BDGP/HHMI AT Drosophila EST Project
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: AT09365.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed. hit
            genomic AE003527: arm:3L [16203031,16482709]
            estimated-cyto:72E3-73A8: 01/30/2002
            Plate: AT.93  row: F  column: 5
            High quality sequence stop: 479.
FEATURES              Location/Qualifiers
     source           1. .610
                      /organism="Drosophila melanogaster"
                      /mol_type="mRNA"
                      /db_xref="taxon:7227"
                      /clone="AT09365"
                      /sex="male"
                      /dev_stage="0-3 day old Ore-R males"
                      /lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
                      AT.121-AT.319: DH5-alpha TonA"
                      /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
                      /note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
                      Site_2: Xho1; The mRNA for the testis library was made
                      from testes and seminal vessicles hand dissected from 0-3
                      day old Ore-R males. RNA kindly provided by the lab of
                      Margaret Fuller. Sized fractionated cDNAs were directly
                      ligated into pOTB7. Plasmid cDNA library."
ORIGIN

  Query Match             8.2%;  Score 37.4;  DB 14;  Length 610;
```

```
  Best Local Similarity   57.1%;  Pred. No. 9.3;
  Matches   68;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        240 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAA 299
              || |   ||||| |   ||  |  ||   |||||||   ||  ||||   ||| |||||
Db        504 TTGTCTGCAGCTTGGGCCCAATTGTGGGTCCCATGATCCGGGCCATGACCTCAAAAATAG 445

Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
              ||||  ||||     | |||| | ||    | | | ||| | | |  | ||| ||| |
Db        444 TGCCCACATCAGAGCGCGGAAAGGTATTTGCACTGCTATCAGTTTGCGATAATGCGGTG 386


RESULT 6
BI667482
LOCUS       BI667482                768 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603292284F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5311557 5',
            mRNA sequence.
ACCESSION   BI667482
VERSION     BI667482.1  GI:15581715
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 768)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11789  row: h  column: 22
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .768
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5311557"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             7.9%;  Score 36.2;  DB 12;  Length 768;
  Best Local Similarity   59.0%;  Pred. No. 23;
  Matches   62;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        163 GCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGT 222
              ||||    ||| |   |||  || |||  |  |    || | | ||   ||||| | ||
Db        427 GCTGATGCTGCCCGAGAAATTGAGGCCGTTGTACAACAATCCAGCAGGTCCCAGAACAGT 486

Qy        223 TATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTT 267
              | |  |   ||| |||||||||| ||   || ||   |||||| |
Db        487 TTTCTTCTGGGCTCCGATTATGAACATGGGGGTAGGTGTGTTGCT 531


RESULT 7
AA263424
LOCUS       AA263424                479 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  LD07239.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD07239 5prime, mRNA sequence.
ACCESSION   AA263424
VERSION     AA263424.1  GI:1899484
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Blast expect value = 2.3E-185 on X55979: Drosophila mRNA for
            vacuolar H+-ATPase Blast expect value = 1.2E-175 on X77936:
            D.melanogaster (Canton S) ductin, subunit C proteolipid gene
            Plate: 72  row: D  column: 3
            High quality sequence stop: 316.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="BDGP_EST:BDcln006624"
                     /db_xref="taxon:7227"
                     /clone="LD07239"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
ORIGIN

  Query Match             7.9%;  Score 36;  DB 9;  Length 479;
  Best Local Similarity   49.5%;  Pred. No. 21;
  Matches   93;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        102 CGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAA 161
              ||  |  ||     || | ||| || | |  | |   ||| |  | |||  ||| ||| ||
Db        175 CGCCATCATCTTCTCGGCCCTGGGCGCTGCTTATGGCACTGCTAAGTCTGGTACCGGTAT 234

Qy        162 CGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGG 221
              |||||  |  |  ||    |  ||  | |       || | | |||| |   | |
Db        235 TGCTGCCATGTCAGTGATGCGGCCTGAACTGATCATGAAATCCATCATTCCTGTGGTCAT 294

Qy        222 TTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGC 281
              ||||  |   |||  ||| | || | |||| ||   || || ||   |  |||
Db        295 GGCGGGTATCATTGCCATTTACGGTCTGGTGGTGGCCGTGCTTATTGCCGGCGCGCTGGA 354

Qy        282 ACATGCAT 289
              | |  |||
Db        355 AGAGCCAT 362


RESULT 8
BI371815
LOCUS       BI371815                545 bp    mRNA    linear   EST 01-AUG-2001
DEFINITION  RE58816.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE58816 5 similar to Vha16:
```

```
            FBan0003161 GO:[hydrogen-transporting ATP synthase (GO:0003936);
            hydrogen-transporting ATP synthase (GO:0003936)] located on: 2R
            42B2-42B2;: 05/16/2001, mRNA sequence.
ACCESSION   BI371815
VERSION     BI371815.1  GI:15067843
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003789: arm:2R [1534085,1833558]
            estimated-cyto:42A13-42C5: 05/16/2001
            Plate: RE.588  row: B  column: 4
            High quality sequence stop: 310.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="RE58816"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
ORIGIN

  Query Match             7.9%;  Score 36;  DB 12;  Length 545;
  Best Local Similarity   49.5%;  Pred. No. 23;
  Matches   93;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        102 CGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAA 161
              ||||  ||      || | || || | |  | |   ||| |  | |||  ||| ||| |
Db        188 CGGCATCATCTTCTCGGCCCTGGGCGCTGCTTATGGCACTGCTAAGTCTGGTACCGGGAT 247

Qy        162 CGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGG 221
               |||||  |  |  ||   | | ||  |  |    ||| | |||| |   |  |
Db        248 TGCTGCCATGTCAGTGATGCGGGCTGAACTGATCATGAAATCCATCATTCCTGTGGTCAT 307

Qy        222 TTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGC 281
                ||||  |   |||  ||| | || | ||||  |      || || ||   |  |||
Db        308 GGCGGGTATCATTGCCATTTACGGTCTGGTGGTGGCTGTGCTTATTGCCGGCGCGCTGGA 367

Qy        282 ACATGCAT 289
              | |  |||
Db        368 AGAGCCAT 375


RESULT 9
BI162493
LOCUS       BI162493                627 bp    mRNA    linear   EST 09-JUL-2001
DEFINITION  RE01790.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE01790 5 similar to Vha16:
```

```
            FBan0003161 'enzyme' located on: 2R 42B2-42B2;: 04/11/2001, mRNA
            sequence.
ACCESSION   BI162493
VERSION     BI162493.1  GI:14628299
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: RE01790.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003789: arm:2R [1534085,1833558]
            estimated-cyto:42A13-42C5: 04/11/2001
            Plate: RE.17  row: H  column: 6
            High quality sequence stop: 609.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="RE01790"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
ORIGIN

  Query Match             7.9%;  Score 36;  DB 12;  Length 627;
  Best Local Similarity   49.5%;  Pred. No. 24;
  Matches   93;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        102 CGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAA 161
              || || ||      || | || || | |  | |   ||| |  | |||  ||| ||| ||
Db        188 CGCCGTCATCTTCTCGGCCCTGGGCGCTGCTTATGGCACTGCTAAGTCTGGTACCGGTAT 247

Qy        162 CGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGG 221
              |||||  |  |  ||   | | ||  |  |    ||| | |||| |   |  |
Db        248 TGCTGCCATGTCAGTGATGCGGCCTGAACTGATCATGAAATCCATCATTCCTGTGGTCAT 307

Qy        222 TTATGGTAACGGCGCCGATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGC 281
                ||||  |   |||  ||| | || | ||||  |      || || ||   |  |||
Db        308 GGCGGGTATCATTGCCATTTACGGTCTGGTGGTGGCTGTGCTTATTGCCGGCGCGCTGGA 367

Qy        282 ACATGCAT 289
              | |  |||
Db        368 AGAGCCAT 375


RESULT 10
BG632414/c
LOCUS       BG632414                543 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  GH08173.3prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH08173 3 similar to CG4526: FBan0004526
```

```
            'transporter' located on: 3L 73A3-73A4;: 04/10/2001, mRNA sequence.
ACCESSION   BG632414
VERSION     BG632414.1  GI:13757599
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: GH08173.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed. hit
            genomic AE003527: arm:3L [16203031,16482709]
            estimated-cyto:72E3-73A8: 04/10/2001
            Plate: GH.81  row: G  column: 1
            High quality sequence stop: 466.
FEATURES                Location/Qualifiers
     source             1. .543
                        /organism="Drosophila melanogaster"
                        /mol_type="mRNA"
                        /db_xref="taxon:7227"
                        /clone="GH08173"
                        /sex="male and female"
                        /dev_stage="adult"
                        /lab_host="DH5 - alpha"
                        /clone_lib="GH Drosophila melanogaster head pOT2"
                        /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                        XhoI; Sized fractionated cDNAs were directly ligated into
                        pOT2.  Plasmid cDNA library."
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 12;  Length 543;
  Best Local Similarity   56.3%;  Pred. No. 26;
  Matches   67;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        240 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAA 299
              || |   ||||| |   ||  |  ||   |||||||   ||  ||||   ||  |||||
Db        504 TTGTCTGCAGCTTGGGCCCAATTGTGGGTCCCATGATCCGGGCCATGACCTCGAAAATAG 445

Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
              ||||  ||||     | ||||  | ||    | | | ||| | | |  | ||| ||| |
Db        444 TGCCCACATCAGAGCGCGGAAAGGTATTTGCACTGCTATCAGTTTGCGATAATGCGGTG 386


RESULT 11
AW944654/c
LOCUS       AW944654                 578 bp    mRNA    linear   EST 02-DEC-2003
DEFINITION  SD10674.3prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD10674 3, mRNA sequence.
ACCESSION   AW944654
VERSION     AW944654.1  GI:8122402
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: SD10674.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed. hit
            genomic AE003527: Drosophila melanogaster genomic scaffold
            142000013386050 section 14 of 54, complete sequence.: 03/20/2001
            Plate: SD.106  row: G  column: 2
            High quality sequence stop: 456.
FEATURES                Location/Qualifiers
     source             1. .578
                        /organism="Drosophila melanogaster"
                        /mol_type="mRNA"
                        /db_xref="taxon:7227"
                        /clone="SD10674"
                        /lab_host="DH5-alpha"
                        /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                        culture pOT2"
                        /note="Vector: pOT2; Site_1: EcoRI; Site_2: Xho1; Sized
                        fractionated cDNAs were directly ligated into pOT2.
                        Plasmid cDNA library."
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 10;  Length 578;
  Best Local Similarity   56.3%;  Pred. No. 26;
  Matches   67;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        240 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAA 299
              || |   ||||| |   ||  |  ||   |||||||   ||  ||||   ||  |||||
Db        494 TTGTCTGCAGCTTGGGCCCAATTGTGGGTCCCATGATCCGGGCCATGACCTCGAAAATAG 435

Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
              ||||  ||||     | ||||  | ||    | | | ||| | | |  | ||| ||| |
Db        434 TGCCCACATCAGAGCGCGGAAAGGTATTTGCACTGCTATCAGTTTGCGATAATGCGGTG 376


RESULT 12
CA806764/c
LOCUS       CA806764                 659 bp    mRNA    linear   EST 09-DEC-2002
DEFINITION  ESG014d.E21_G05.3prime ESG01 Drosophila melanogaster cDNA 3'
            similar to CT43205, mRNA sequence.
ACCESSION   CA806764
VERSION     CA806764.1  GI:26255713
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Gorski,S.M., Chittaranjan,S., Pleasance,E.D., Freeman,J.D.,
            Anderson,C.L., Varhol,R.J., Coughlin,S.M., Zuyderduyn,S.D.,
            Jones,S.J.M. and Marra,M.A.
  TITLE     A SAGE Approach to Discovery of Genes Involved in Autophagic Cell
            Death
  JOURNAL   Curr. Biol. 13 (4), 358-363 (2003)
  MEDLINE   22482219
   PUBMED   12593804
COMMENT     Contact: Gorski SM
            Genome Sciences Centre
            British Columbia Cancer Agency
            600 West 10th Ave., Room 3427, Vancouver, BC, Canada, V5Z 4E6
            Tel: 604 877 6081
            Fax: 604 877 6085
            Email: sgorski@bcgsc.bc.ca
            Seq primer: -21M13F.
FEATURES                Location/Qualifiers
```

```
     source          1. .659
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /strain="Oregon R"
                     /db_xref="taxon:7227"
                     /sex="male and female"
                     /tissue_type="salivary gland"
                     /dev_stage="16, 18, 20, 22, and 24 hrs after puparium
                     formation at 18oC"
                     /lab_host="DH10B"
                     /clone_lib="ESG01"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; mRNA
                     isolated from 500 pairs of hand-dissected salivary glands
                     from mixed stage animals was used to construct an
                     oligo-dT-primed directional cDNA library"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 14;  Length 659;
  Best Local Similarity   56.3%;  Pred. No. 28;
  Matches   67;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        240 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAA 299
              || |   ||||| |   ||  |  ||   |||||||   ||  ||||   ||  |||||
Db        512 TTGTCTGCAGCTTGGGCCCAATTGTGGGTCCCATGATCCGGGCCATGACCTCGAAAATAG 453

Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
              ||||  ||||     | |||| | ||    | | | ||| | | |  |  ||| ||| |
Db        452 TGCCCACATCAGAGCGCGGAAAGGTATTTGCACTGCTATCAGTTTGCGATAATGCGGTG 394


RESULT 13
CA804002/c
LOCUS       CA804002                678 bp    mRNA    linear   EST 09-DEC-2002
DEFINITION  ESG0111c.B21_G04.3prime ESG01 Drosophila melanogaster cDNA 3'
            similar to CT43205, mRNA sequence.
ACCESSION   CA804002
VERSION     CA804002.1  GI:26252951
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Gorski,S.M., Chittaranjan,S., Pleasance,E.D., Freeman,J.D.,
            Anderson,C.L., Varhol,R.J., Coughlin,S.M., Zuyderduyn,S.D.,
            Jones,S.J.M. and Marra,M.A.
  TITLE     A SAGE Approach to Discovery of Genes Involved in Autophagic Cell
            Death
  JOURNAL   Curr. Biol. 13 (4), 358-363 (2003)
  MEDLINE   22482219
   PUBMED   12593804
COMMENT     Contact: Gorski SM
            Genome Sciences Centre
            British Columbia Cancer Agency
            600 West 10th Ave., Room 3427, Vancouver, BC, Canada, V5Z 4E6
            Tel: 604 877 6081
            Fax: 604 877 6085
            Email: sgorski@bcgsc.bc.ca
            Seq primer: -21M13F.
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /strain="Oregon R"
                     /db_xref="taxon:7227"
                     /sex="male and female"
                     /tissue_type="salivary gland"
                     /dev_stage="16, 18, 20, 22, and 24 hrs after puparium
                     formation at 18oC"
                     /lab_host="DH10B"
                     /clone_lib="ESG01"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; mRNA
                     isolated from 500 pairs of hand-dissected salivary glands
                     from mixed stage animals was used to construct an
                     oligo-dT-primed directional cDNA library"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 14;  Length 678;
  Best Local Similarity   56.3%;  Pred. No. 28;
  Matches   67;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        240 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAA 299
              || |   ||||| |   ||  |  ||   |||||||   ||  ||||   ||  |||||
Db        505 TTGTCTGCAGCTTGGGCCCAATTGTGGGTCCCATGATCCGGGCCATGACCTCGAAAATAG 446

Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
              ||||  ||||     | |||| | ||    | | | ||| | | |  |  ||| ||| |
Db        445 TGCCCACATCAGAGCGCGGAAAGGTATTTGCACTGCTATCAGTTTGCGATAATGCGGTG 387


RESULT 14
CA807388/c
LOCUS       CA807388                723 bp    mRNA    linear   EST 09-DEC-2002
DEFINITION  ESG017c.E21_H06.3prime ESG01 Drosophila melanogaster cDNA 3'
            similar to CT43205, mRNA sequence.
ACCESSION   CA807388
VERSION     CA807388.1  GI:26256337
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Gorski,S.M., Chittaranjan,S., Pleasance,E.D., Freeman,J.D.,
            Anderson,C.L., Varhol,R.J., Coughlin,S.M., Zuyderduyn,S.D.,
            Jones,S.J.M. and Marra,M.A.
  TITLE     A SAGE Approach to Discovery of Genes Involved in Autophagic Cell
            Death
  JOURNAL   Curr. Biol. 13 (4), 358-363 (2003)
  MEDLINE   22482219
   PUBMED   12593804
COMMENT     Contact: Gorski SM
            Genome Sciences Centre
            British Columbia Cancer Agency
            600 West 10th Ave., Room 3427, Vancouver, BC, Canada, V5Z 4E6
            Tel: 604 877 6081
            Fax: 604 877 6085
            Email: sgorski@bcgsc.bc.ca
            Seq primer: -21M13F.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /strain="Oregon R"
                     /db_xref="taxon:7227"
                     /sex="male and female"
                     /tissue_type="salivary gland"
                     /dev_stage="16, 18, 20, 22, and 24 hrs after puparium
                     formation at 18oC"
                     /lab_host="DH10B"
                     /clone_lib="ESG01"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; mRNA
                     isolated from 500 pairs of hand-dissected salivary glands
                     from mixed stage animals was used to construct an
                     oligo-dT-primed directional cDNA library"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 14;  Length 723;
  Best Local Similarity   56.3%;  Pred. No. 29;
  Matches   67;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        240 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAA 299
              || |   ||||| |   ||  |  ||   |||||||   ||  ||||   ||  |||||
Db        519 TTGTCTGCAGCTTGGGCCCAATTGTGGGTCCCATGATCCGGGCCATGACCTCGAAAATAG 460
```

```
Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
              ||||  ||||     | |||| | ||    | | | ||| | | |  |  ||| ||| |
Db        459 TGCCCACATCAGAGCGCGGAAAGGTATTTGCACTGCTATCAGTTTGCGATAATGCGGTG 401


RESULT 15
BG633139/c
LOCUS       BG633139                800 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  GH20822.3prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH20822 3 similar to CG4526: FBan0004526
            'transporter' located on: 3L 73A3-73A4;: 04/10/2001, mRNA sequence.
ACCESSION   BG633139
VERSION     BG633139.1  GI:13758766
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: GH20822.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed. hit
            genomic AE003527: arm:3L [16203031,16482709]
            estimated-cyto:72E3-73A8: 04/10/2001
            Plate: GH.208  row: B  column: 10
            High quality sequence stop: 720.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="GH20822"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 12;  Length 800;
  Best Local Similarity   56.3%;  Pred. No. 31;
  Matches   67;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        240 TTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAA 299
              || |   ||||| |   ||  |  ||  |||||||   ||  ||||   ||  |||||
Db        430 TTGTCTGCAGCTTGGGCCCAATTGTGGGTCCCATGATCCGGGCCATGACCTCGAAAATAG 371

Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCG 358
              ||||  ||||     | |||| | ||    | | | ||| | | |  |  ||| ||| |
Db        370 TGCCCACATCAGAGCGCGGAAAGGTATTTGCACTGCTATCAGTTTGCGATAATGCGGTG 312


Search completed: March 15, 2004, 10:09:31
Job time : 1966.4 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw_model

Run on: March 14, 2004, 19:02:18 ; Search time 2269.8 Seconds
(without alignments)
8707.562 Million cell updates/sec

Title: US-09-543-407-29
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		398.4	87.4	456	6	I44909	I44909 Sequence 58
2		398.4	87.4	2067	1	SEU43280	U43280 Salmonella
3		392	86.0	5103	1	STAJ2301	AJ002301 Salmonell
4		392	86.0	22411	1	AE008749	AE008749 Salmonell
5		390.4	85.6	254050	1	AL627269	AL627269 Salmonell
6	c	390.4	85.6	301983	1	AE016840	AE016840 Salmonell
7		384.6	84.3	1048	1	STAGFBA	AJ000514 Salmonell
8		274.2	60.1	361	6	I44908	I44908 Sequence 56
9		273.6	60.0	2889	1	CSP515700	AJ515700 Citrobact
10		235.6	51.7	2920	1	CFR515701	AJ515701 Citrobact
11		228.8	50.2	4680	1	ECCSGABDG	X90754 E.coli csgG
12		228.8	50.2	10346	1	AE000205	AE000205 Escherich
13		228.8	50.2	15047	1	D90741	D90741 Escherichia
14		227.2	49.8	456	6	AX814811	AX814811 Sequence
15		227.2	49.8	648	1	ECOCSGAA	L04979 Escherichia
16		227	49.8	306358	1	AE016759	AE016759 Escherich
17		220.6	48.4	230	1	SEU53207	U53207 Salmonella
18		217.4	47.7	1711	1	AF275733	AF275733 Escherich
19		217.4	47.7	10190	1	AE005315	AE005315 Escherich
20		217.4	47.7	327773	1	AP002554	AP002554 Escherich
21		185.6	40.7	2883	1	ESA515702	AJ515702 Enterobac
22		161.6	35.4	10370	1	AE015131	AE015131 Shigella
23		161.6	35.4	292504	1	AE016981	AE016981 Shigella
24	c	138.6	30.4	437	1	AF237726	AF237726 Shigella
25		66	14.5	19201	1	D90742	D90742 Escherichia
26		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27		48.2	10.6	78	6	AX814809	AX814809 Sequence
28		44.8	9.8	72	6	AX814798	AX814798 Sequence
29	c	39.8	8.7	302156	3	AC116977	AC116977 Dictyoste
30	c	38.8	8.5	2000	6	AX655393	AX655393 Sequence
31	c	37.6	8.2	233367	10	AC118931	AC118931 Mus muscu
32		36.8	8.1	336382	2	AC127747	AC127747 Rattus no
33	c	36	7.9	168267	2	AC146009	AC146009 Pan trogl
34	c	35.6	7.8	82727	1	ECO82K	D26562 Escherichia
35		35.2	7.7	194439	3	CNS07EGA	AL590443 chromosom
36	c	35.2	7.7	205317	9	AC093856	AC093856 Homo sapi
37	c	35	7.7	1425	1	AF343445	AF343445 Lactobaci
38	c	34.8	7.6	7366	1	ECH132326	AJ132326 Erwinia c
39		34.4	7.5	11160	1	AE008209	AE008209 Agrobacte
40	c	34.4	7.5	11476	1	AE009405	AE009405 Agrobacte
41		34.4	7.5	151828	2	AC146008	AC146008 Pan trogl
42	c	34.4	7.5	156596	2	AL355479	AL355479 Homo sapi
43	c	34	7.5	1344	6	BD243128	BD243128 Hypersens
44	c	34	7.5	1344	6	BD243666	BD243666 Hypersens
45	c	34	7.5	1344	6	AR403552	AR403552 Sequence

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             87.4%;  Score 398.4;  DB 6;  Length 456;
  Best Local Similarity   92.1%;  Pred. No. 5.3e-99;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |   |          |  | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match             87.4%;  Score 398.4;  DB 1;  Length 2067;
  Best Local Similarity   92.1%;  Pred. No. 5.3e-99;
```

```
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |   |           |  | |  ||    ||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN
```

Query Match 86.0%; Score 392; DB 1; Length 5103;
Best Local Similarity 91.2%; Pred. No. 3.1e-97;
Matches 416; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |  |                |   | | ||    ||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421
```

```
RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
     RBS                complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene               complement(2580. .4088)
                        /gene="STM1128"
     CDS                complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS                complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene               complement(4413. .5099)
                        /gene="STM1129"
     CDS                complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
     RBS                complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene               5599. .6759
                        /gene="STM1130"
     CDS                5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene               6793. .7497
                        /gene="STM1131"
     RBS                6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match             86.0%;  Score 392;  DB 1;  Length 22411;
  Best Local Similarity   91.2%;  Pred. No. 3.1e-97;
  Matches  416;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |    |                |    | |  ||    ||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAININSSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAARERKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             85.6%;  Score 390.4;  DB 1;  Length 254050;
  Best Local Similarity   91.0%;  Pred. No. 8.4e-97;
  Matches  415;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| ||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |   |           |   | |  ||    ||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361
```

RESULT 6
AE016840/c
LOCUS AE016840 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of 16 of the complete genome.
ACCESSION AE016840 AE014613
VERSION AE016840.1 GI:29137797
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi Ty2
ORGANISM Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 301983)
AUTHORS Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and CT18
JOURNAL J. Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE 22531367
PUBMED 12644504
REFERENCE 2 (bases 1 to 301983)
AUTHORS Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES Location/Qualifiers

```
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match           85.6%;  Score 390.4;  DB 1;  Length 301983;
Best Local Similarity 91.0%;  Pred. No. 8.4e-97;
  Matches  415;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || ||| |  |           |  | |  ||    ||||||
Db.     37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             84.3%;  Score 384.6;  DB 1;  Length 1048;
  Best Local Similarity   90.3%;  Pred. No. 3.4e-95;
  Matches  411;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |   |               |   | |  ||    ||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              ||||||||||||||||  |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             60.1%;  Score 274.2;  DB 6;  Length 361;
  Best Local Similarity   88.7%;  Pred. No. 1.1e-64;
  Matches  297;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAAT 363
              |||   || |||  |   |                 | | ||    |||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           60.0%;  Score 273.6;  DB 1;  Length 2889;
  Best Local Similarity  76.5%;  Pred. No. 1.6e-64;
  Matches  349;  Conservative    0;  Mismatches  104;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||    |     ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||    |  |||||| ||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||    ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   ||||||  |      |      |      | |    ||   ||    || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || ||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
CFR515701
LOCUS       CFR515701              2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
  Query Match             51.7%;  Score 235.6;  DB 1;  Length 2920;
  Best Local Similarity   71.5%;  Pred. No. 5e-54;
  Matches  326;  Conservative    0;  Mismatches  124;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||    ||| ||||   ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||    ||||| || ||   |||  ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||||| ||| | ||||| |||   ||||  |||| ||  |||  || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              |||||     ||||||  |  | ||       || |      |    || | ||   || ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||| ||||||||||||||| || ||||| ||||  ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572


RESULT 11
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
```

```
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           50.2%;  Score 228.8;  DB 1;  Length 4680;
  Best Local Similarity   68.9%;  Pred. No. 3.8e-52;
  Matches  314;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || |||| || |  |||||||||||||| ||    || |||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| ||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              || ||  |||||||  |      |      | |          |   |||    || |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 12
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
```

and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

```
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423
```

```
Query Match           50.2%;  Score 228.8;  DB 1;  Length 10346;
Best Local Similarity  68.9%;  Pred. No. 3.8e-52;
Matches  314;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              || ||  |||||||  |      |      | |           |   |||    || |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

RESULT 13
D90741
LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai;K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
FEATURES Location/Qualifiers
source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDMEDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHEPVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNSSFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPPER

```
                ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS             2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
CDS             3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
gene            complement(4117. .4950)
                /gene="csgG"
CDS             complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
gene            complement(4977. .5393)
                /gene="csgF"
CDS             complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene            complement(5418. .5807)
                /gene="csgE"
CDS             complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
gene            complement(5812. .6462)
                /gene="csgD"
CDS             complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene            7217. .7672
                /gene="csgB"
CDS             7217. .7672
```

Query Match 50.2%; Score 228.8; DB 1; Length 15047;
Best Local Similarity 68.9%; Pred. No. 3.8e-52;
Matches 314; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||    | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||||  || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              || ||   |||||||  |      |  |  |           |  |||    || |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168
```

```
RESULT 14
AX814811
LOCUS       AX814811               456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
```

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           49.8%;  Score 227.2;  DB 6;  Length 456;
  Best Local Similarity  68.6%;  Pred. No. 1.1e-51;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | ||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||||  || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              || ||   |||||||  |     |     | |           |   |||    || |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match           49.8%;  Score 227.2;  DB 1;  Length 648;
  Best Local Similarity  68.6%;  Pred. No. 1.1e-51;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | ||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||||  || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              || ||   |||||||  |     |     | |           |   |||    || |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 15, 2004, 07:24:18
Job time : 2273.2 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2004, 18:52:53 ; Search time 290.4 Seconds
(without alignments)
6670.724 Million cell updates/sec

Title: US-09-543-407-29
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64631	Aac64631 AgfA::PT3
	2	398.4	87.4	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	398.4	87.4	456	2	AAT74142	Aat74142 Salmonell
	4	398.4	87.4	456	3	AAC64617	Aac64617 Salmonell
	5	377.6	82.8	456	3	AAC64628	Aac64628 AgfA::PT3
	6	355.2	77.9	456	3	AAC64626	Aac64626 AgfA::PT3
	7	348.8	76.5	456	3	AAC64625	Aac64625 AgfA::PT3
	8	348.8	76.5	456	3	AAC64623	Aac64623 AgfA::PT3
	9	348.8	76.5	456	3	AAC64627	Aac64627 AgfA::PT3
	10	347.2	76.1	456	3	AAC64622	Aac64622 AgfA::PT3
	11	347.2	76.1	456	3	AAC64629	Aac64629 AgfA::PT3
	12	345.6	75.8	456	3	AAC64624	Aac64624 AgfA::PT3
	13	344	75.4	456	3	AAC64630	Aac64630 AgfA::PT3
	14	274.2	60.1	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	274.2	60.1	361	2	AAT74141	Aat74141 Salmonell
	16	227.2	49.8	456	3	AAC64619	Aac64619 Escherich
	17	227.2	49.8	456	9	ACF36153	Acf36153 E. coli C
	18	214.8	47.1	646	2	AAQ62647	Aaq62647 Fibronect
	19	152.8	33.5	369	2	AAQ62646	Aaq62646 FNB curli
c	20	49.4	10.8	78	3	AAC64609	Aac64609 AgfA (SEF
	21	49	10.7	100	7	ACD68807	Acd68807 E. coli K
	22	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	23	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	24	48	10.5	78	3	AAC64610	Aac64610 AgfA (SEF
	25	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
	26	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	27	42.6	9.3	78	3	AAC64606	Aac64606 SefA (SEF
	28	42.4	9.3	100	7	ACD68809	Acd68809 E. coli K
c	29	41	9.0	78	3	AAC64605	Aac64605 SefA (SEF
c	30	38.8	8.5	2000	7	ADA71938	Ada71938 Rice gene
	31	35	7.7	100	7	ACD68808	Acd68808 E. coli K
	32	34.8	7.6	1083	5	AAS76745	Aas76745 DNA encod
c	33	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	34	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	35	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	36	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	37	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	38	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c	39	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a
c	40	33.6	7.4	388	7	ABX42984	Abx42984 Bovine ES
	41	33.6	7.4	2649	2	AAQ51026	Aaq51026 Human glu
	42	33.6	7.4	2752	5	AAS06006	Aas06006 Angiotens
	43	33.6	7.4	2955	2	AAT02798	Aat02798 Human glu
	44	33.6	7.4	2955	2	AAT02797	Aat02797 Human glu
	45	33.6	7.4	3331	7	ACC50172	Acc50172 Breast ca

ALIGNMENTS

RESULT 1
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36355.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

```
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-137;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag=  a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match           87.4%;  Score 398.4;  DB 2;  Length 456;
  Best Local Similarity  92.1%;  Pred. No. 6.3e-119;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |  |                  |   | |  ||  ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
AAT74142
```

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             87.4%;  Score 398.4;  DB 2;  Length 456;
  Best Local Similarity   92.1%;  Pred. No. 6.3e-119;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |   |               |   | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

```
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             87.4%;  Score 398.4;  DB 3;  Length 456;
  Best Local Similarity   92.1%;  Pred. No. 6.3e-119;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |    |          |   | |  ||   ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             82.8%;  Score 377.6;  DB 3;  Length 456;
  Best Local Similarity   89.3%;  Pred. No. 3.6e-112;
  Matches  407;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |    |          |   ||     |  | | ||  |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |   || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 6
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 6.9e-105;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| |||    |  ||||||    | |    | |   ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |    |                |   | | ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 7
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 8.3e-103;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |    ||   |      | | ||   | |     || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |   |                |   | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match             76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 8.3e-103;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |   |                |   | | ||   |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||   | ||||  ||       |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match             76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 8.3e-103;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |  |      | ||  |    ||  |   |  | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||       || |||  |   |              |   | |  ||    ||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

```
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             76.1%;  Score 347.2;  DB 3;  Length 456;
  Best Local Similarity   85.1%;  Pred. No. 2.8e-102;
  Matches  388;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              |||||||  || |||  |   |                |  | |  ||   |||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||| |           | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
```

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match            76.1%;  Score 347.2;  DB 3;  Length 456;
  Best Local Similarity  85.1%;  Pred. No. 2.8e-102;
  Matches  388;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||  |        ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |    |                |    | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64624
ID   AAC64624 standard; DNA; 456 BP.
XX
AC   AAC64624;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36348.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

  Query Match            75.8%;  Score 345.6;  DB 3;  Length 456;
  Best Local Similarity  84.9%;  Pred. No. 9.1e-102;
  Matches  387;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |    ||   ||     | | | |    | |      |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |    |                |    | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64630
ID   AAC64630 standard; DNA; 456 BP.
XX
AC   AAC64630;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36354.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match             75.4%;  Score 344;  DB 3;  Length 456;
  Best Local Similarity   84.6%;  Pred. No. 3e-101;
  Matches  386;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||| |  | | | || |  |  |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |  |  |  |  | ||  ||  ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             60.1%;  Score 274.2;  DB 2;  Length 361;
  Best Local Similarity   88.7%;  Pred. No. 1.3e-78;
  Matches  297;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAAT 363
              |||   || |||  |    |               |   | |  ||    |||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             60.1%;  Score 274.2;  DB 2;  Length 361;
  Best Local Similarity   88.7%;  Pred. No. 1.3e-78;
  Matches  297;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAAT 363
              |||   || |||  |    |               |   | |  ||    |||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 15, 2004, 04:14:46
```

Job time : 291.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:01:03 ; Search time 56.8 Seconds
(without alignments)
4455.240 Million cell updates/sec

Title: US-09-543-407-29
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	398.4	87.4	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	274.2	60.1	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	33.6	7.4	2649	4	US-08-400-067-2	Sequence 2, Appli
	6	33.6	7.4	2955	1	US-08-687-379-5	Sequence 5, Appli
	7	33.6	7.4	2955	1	US-08-687-379-7	Sequence 7, Appli
	8	33.6	7.4	3407	3	US-08-483-327-1	Sequence 1, Appli
	9	33.6	7.4	3407	3	US-08-666-221B-1	Sequence 1, Appli
	10	33.6	7.4	3407	3	US-08-666-221B-7	Sequence 7, Appli
	11	33.6	7.4	3407	4	US-08-439-946-1	Sequence 1, Appli
	12	33.6	7.4	3407	4	US-08-242-344-1	Sequence 1, Appli
	13	32.6	7.1	129908	4	US-09-585-858-1	Sequence 1, Appli
	14	32.4	7.1	2166	2	US-08-317-401E-3	Sequence 3, Appli
	15	32.4	7.1	2877	2	US-08-317-401E-1	Sequence 1, Appli
	16	31.8	7.0	827	4	US-08-976-259-36	Sequence 36, Appl
	17	31.8	7.0	1613	4	US-08-976-259-111	Sequence 111, App
	18	30.8	6.8	906	4	US-09-134-001C-383	Sequence 383, App
c	19	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	20	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	21	30.2	6.6	5976	4	US-08-621-944A-2	Sequence 2, Appli
	22	30.2	6.6	5976	4	US-08-945-567D-2	Sequence 2, Appli
	23	30.2	6.6	6973	1	US-08-478-370-1	Sequence 1, Appli
	24	30.2	6.6	6973	4	US-08-483-855C-1	Sequence 1, Appli
	25	30.2	6.6	6973	4	US-08-621-944A-1	Sequence 1, Appli
	26	30.2	6.6	6973	4	US-08-945-567D-1	Sequence 1, Appli
	27	30.2	6.6	6975	4	US-08-431-718C-1	Sequence 1, Appli
c	28	30.2	6.6	10411	4	US-08-961-527-89	Sequence 89, Appl
	29	30.2	6.6	16995	4	US-08-961-527-82	Sequence 82, Appl
c	30	29.8	6.5	837	4	US-09-489-039A-1827	Sequence 1827, Ap
	31	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	32	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	33	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
	34	29.6	6.5	1047	4	US-09-445-803-1	Sequence 1, Appli
	35	29.6	6.5	7766	4	US-09-125-619-3	Sequence 3, Appli
c	36	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
	37	29.4	6.4	786	4	US-09-134-000C-1437	Sequence 1437, Ap
c	38	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	39	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	40	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
c	41	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
	42	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
	43	29.2	6.4	51259	3	US-08-781-891-209	Sequence 209, App
	44	29.2	6.4	51259	4	US-09-618-166-209	Sequence 209, App
	45	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 87.4%; Score 398.4; DB 1; Length 456;
Best Local Similarity 92.1%; Pred. No. 3.3e-124;
Matches 420; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || ||| |   |                |  | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 60.1%; Score 274.2; DB 1; Length 361;
Best Local Similarity 88.7%; Pred. No. 1.8e-82;
Matches 297; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAAT 363
              |||   || |||  |   |                |    | |  ||    |||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA AND ITS USE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,927
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,108
; FILING DATE: 06-AUG-1977
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1581
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304

```
;       TELEFAX:  (716) 263-1600
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1344 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.22;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.22;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-08-400-067-2
; Sequence 2, Application US/08400067
; Patent No. 6312945
;  GENERAL INFORMATION:
;    APPLICANT:  Burnett, J. Paul
;    APPLICANT:  Mayne, Nancy G.
;    APPLICANT:  Sharp, Robert L.
;    APPLICANT:  Snyder, Yvonne M.
;    TITLE OF INVENTION:  Human Glutamate Receptor Proteins      and
;    TITLE OF INVENTION:  Associated DNA Compounds
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company/Patent Division
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/400,067
;      FILING DATE:  March 7, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Blalock, Donna K.
;      REGISTRATION NUMBER:  38,082
;      REFERENCE/DOCKET NUMBER:  X-8870B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317/277-1090
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2649 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-400-067-2

  Query Match             7.4%;  Score 33.6;  DB 4;  Length 2649;
  Best Local Similarity   49.4%;  Pred. No. 0.42;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |      |||||   | | || | ||  |  ||    | |  ||
Db       1117 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1176

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              | |     |||||||||| |     |  | | | ||  | | | | | | ||| |
Db       1177 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1236

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   | |||    || |  || | |    || ||
Db       1237 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1292


RESULT 6
US-08-687-379-5
; Sequence 5, Application US/08687379
; Patent No. 5756697
;  GENERAL INFORMATION:
;    APPLICANT:  Hoeger, Thomas
;    APPLICANT:  Ultsch, Andreas
;    APPLICANT:  Bach, Alfred
;    APPLICANT:  Sterrer, Sylvia
;    APPLICANT:  Lemaire, Hans-Georg
;    TITLE OF INVENTION:  Subunits of Glutamate Receptors, Their
;    TITLE OF INVENTION:  Preparation and Their Use
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Keil & Weinkauf
;      STREET:  1101 Connecticut Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 720 Kb storage
;      COMPUTER:  IBM AT-compatible, 80286 processor
;      OPERATING SYSTEM:  MS-DOS version 6.0
;      SOFTWARE:  WordPerfect version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,379
;      FILING DATE:  05-AUG-1996
;      CLASSIFICATION:  435
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2955 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      DEVELOPMENTAL STAGE:  Adult
;      TISSUE TYPE:  Brain
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..2676
US-08-687-379-5

  Query Match             7.4%;  Score 33.6;  DB 1;  Length 2955;
  Best Local Similarity   49.4%;  Pred. No. 0.44;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||  | | || | || |  ||    | |  ||
Db       1144 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1203

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |  | | | ||  | | | | | | |||  |
Db       1204 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1263

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   | |||    || |  || | |    || ||
Db       1264 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1319


RESULT 7
US-08-687-379-7
; Sequence 7, Application US/08687379
; Patent No. 5756697
;  GENERAL INFORMATION:
;    APPLICANT:  Hoeger, Thomas
;    APPLICANT:  Ultsch, Andreas
;    APPLICANT:  Bach, Alfred
;    APPLICANT:  Sterrer, Sylvia
;    APPLICANT:  Lemaire, Hans-Georg
;    TITLE OF INVENTION:  Subunits of Glutamate Receptors, Their
;    TITLE OF INVENTION:  Preparation and Their Use
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Keil & Weinkauf
;      STREET:  1101 Connecticut Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 720 Kb storage
;      COMPUTER:  IBM AT-compatible, 80286 processor
;      OPERATING SYSTEM:  MS-DOS version 6.0
;      SOFTWARE:  WordPerfect version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,379
;      FILING DATE:  05-AUG-1996
;      CLASSIFICATION:  435
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2955 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      DEVELOPMENTAL STAGE:  Adult
;      TISSUE TYPE:  Brain
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..2676
US-08-687-379-7

  Query Match             7.4%;  Score 33.6;  DB 1;  Length 2955;
  Best Local Similarity   49.4%;  Pred. No. 0.44;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||  | | || | || |  ||    | |  ||
Db       1144 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1203

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |  | | | ||  | | | | | | |||  |
Db       1204 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1263

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   | |||    || |  || | |    || ||
Db       1264 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1319


RESULT 8
US-08-483-327-1
; Sequence 1, Application US/08483327
; Patent No. 6040175
;  GENERAL INFORMATION:
;    APPLICANT:  KAMBOJ, Rajender
;    APPLICANT:  ELIOTT, Candace
;    APPLICANT:  NUTT, Stephen
;    TITLE OF INVENTION:  AMPA-BINDING HUMAN GluR2 RECEPTORS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  Suite 500, 3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,327
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/242,344
;      FILING DATE:  13-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  16777/275/ALLE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3407 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  315..2966
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  315..374
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  375..2966
US-08-483-327-1

  Query Match             7.4%;  Score 33.6;  DB 3;  Length 3407;
  Best Local Similarity   49.4%;  Pred. No. 0.47;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||  | | || | || |  ||    | |  ||
Db       1431 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1490

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |  | | | ||  | | | | | | ||| |
Db       1491 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1550

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||  |||  | |||    || |  || | |    || ||
Db       1551 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1606


RESULT 9
US-08-666-221B-1
; Sequence 1, Application US/08666221B
; Patent No. 6136544
;  GENERAL INFORMATION:
;    APPLICANT:  Kamboj, Rajender
;    APPLICANT:  Nutt, Stephen
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR (OR EAA RECEPTOR)
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND THEIR USES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/666,221B
;      FILING DATE:  20-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  016777/0308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3407 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  315..2966
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  315..374
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  375..2966
US-08-666-221B-1

  Query Match             7.4%;  Score 33.6;  DB 3;  Length 3407;
  Best Local Similarity   49.4%;  Pred. No. 0.47;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||  | | || | || |  ||    | |  ||
Db       1431 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1490

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |  | | | ||  | | | | | | ||| |
Db       1491 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1550

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||  |||  | |||    || |  || | |    || ||
Db       1551 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1606


RESULT 10
US-08-666-221B-7
; Sequence 7, Application US/08666221B
; Patent No. 6136544
;  GENERAL INFORMATION:
;    APPLICANT:  Kamboj, Rajender
;    APPLICANT:  Nutt, Stephen
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR (OR EAA RECEPTOR)
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND THEIR USES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/666,221B
;      FILING DATE:  20-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  016777/0308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3407 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
```

```
;      NAME/KEY:  CDS
;      LOCATION:  315..2966
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  315..374
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  375..2966
US-08-666-221B-7

  Query Match             7.4%;  Score 33.6;  DB 3;  Length 3407;
  Best Local Similarity   49.4%;  Pred. No. 0.47;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||  | | || | || |  ||    | |  ||
Db       1431 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1490

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |  | | ||  | | | | | | ||| |
Db       1491 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1550

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   | |||    || |  || | |    || ||
Db       1551 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1606


RESULT 11
US-08-439-946-1
; Sequence 1, Application US/08439946
; Patent No. 6399313
;  GENERAL INFORMATION:
;    APPLICANT:  KAMBOJ, Rajender
;    APPLICANT:  ELIOTT, Candace
;    APPLICANT:  NUTT, Stephen
;    TITLE OF INVENTION:  AMPA-BINDING HUMAN GluR2 RECEPTORS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  Suite 500, 3000 K Street, N.W.
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,946
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/091,437
;      FILING DATE:  15-JUL-1993
;      APPLICATION NUMBER:  US 07/896,437
;      FILING DATE:  10-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  16777/217 ALLE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3407 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  315..2966
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  315..374
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  375..2966
US-08-439-946-1

  Query Match             7.4%;  Score 33.6;  DB 4;  Length 3407;
  Best Local Similarity   49.4%;  Pred. No. 0.47;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||  | | || | || |  ||    | |  ||
Db       1431 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1490

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |  | | ||  | | | | | | ||| |
Db       1491 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1550

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   | |||    || |  || | |    || ||
Db       1551 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1606


RESULT 12
US-08-242-344-1
; Sequence 1, Application US/08242344
; Patent No. 6441155
;  GENERAL INFORMATION:
;    APPLICANT:  KAMBOJ, Rajender
;    APPLICANT:  ELIOTT, Candace
;    APPLICANT:  NUTT, Stephen
;    TITLE OF INVENTION:  AMPA-BINDING HUMAN GluR2 RECEPTORS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  Suite 500, 3000 K Street, N.W.
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/242,344
;      FILING DATE:  13-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  16777/180/ALLE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3407 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  315..2966
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  315..374
;    FEATURE:
```

```
;      NAME/KEY:  mat_peptide
;      LOCATION:  375..2966
US-08-242-344-1

  Query Match             7.4%;  Score 33.6;  DB 4;  Length 3407;
  Best Local Similarity   49.4%;  Pred. No. 0.47;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||   | | || | || |  ||    | |  ||
Db       1431 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1490

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |      |   | | | ||   | | | | | | ||| |
Db       1491 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1550

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   | |||    || |  || | |    || ||
Db       1551 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1606


RESULT 13
US-09-585-858-1
; Sequence 1, Application US/09585858
; Patent No. 6492161
; GENERAL INFORMATION:
;  APPLICANT:  Sigridur Hjorleifsdotter
;  APPLICANT:  Gudmundur O. Hreggvidsson
;  APPLICANT:  Olafur H. Fridjonsson
;  APPLICANT:  Arnthor Aevarsson
;  APPLICANT:  Jakob K. Kristjansson
;  TITLE OF INVENTION:  Bacteriophage RM378 of a Thermophilic
;  TITLE OF INVENTION:  Host Organism
;  FILE REFERENCE:  2739.1001-001
;  CURRENT APPLICATION NUMBER: US/09/585,858
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER:  60/137,120
;  PRIOR FILING DATE:  1999-06-02
;  NUMBER OF SEQ ID NOS:  73
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 129908
;   TYPE: DNA
;   ORGANISM: Bacteriophage RM378
US-09-585-858-1

  Query Match             7.1%;  Score 32.6;  DB 4;  Length 129908;
  Best Local Similarity   53.5%;  Pred. No. 6.5;
  Matches   68;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        196 TCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCG 255
              | |||||  |   |  |   | | || ||||||  |  |      || | |  ||||| |||
Db      99768 TATGAAAGGGTGACAAGGTATACCGCTTATGGAATGGCAATTAATCTGGTACAGGGGGCG 99827

Qy        256 GATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAG 315
              |  |  |    ||||||   ||| |  | |   |||||  ||   |  |  ||  | | |
Db      99828 CTTGAACGCTACATTGAAACTACTGATGAAGTGATCAGAGTTACCGGTATTTACAACCGG 99887

Qy        316 CTGGTTA 322
              ||| |||
Db      99888 CTGCTTA 99894


RESULT 14
US-08-317-401E-3
; Sequence 3, Application US/08317401E
; Patent No. 5922561
;  GENERAL INFORMATION:
;    APPLICANT:  Thompson, Sheryl Ann
;    APPLICANT:  Yaver, Debbie Sue
;    TITLE OF INVENTION:  GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
;    TITLE OF INVENTION:  ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,401E
;      FILING DATE:  03-October-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4248.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2166 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
US-08-317-401E-3

  Query Match             7.1%;  Score 32.4;  DB 2;  Length 2166;
  Best Local Similarity   68.2%;  Pred. No. 0.95;
  Matches   45;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        275 TGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTA 334
              |||   ||||  |||| |   | || |||  ||  ||| |||||||| |||  |||| |
Db        291 TGAACAAGAAGCGTTTGATTCAAAAGGCCGTCTTCGATGAGCTGGTTTCCCTGGTTGATC 350

Qy        335 CCCATG 340
              ||||||
Db        351 CCCATG 356


RESULT 15
US-08-317-401E-1
; Sequence 1, Application US/08317401E
; Patent No. 5922561
;  GENERAL INFORMATION:
;    APPLICANT:  Thompson, Sheryl Ann
;    APPLICANT:  Yaver, Debbie Sue
;    TITLE OF INVENTION:  GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
;    TITLE OF INVENTION:  ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,401E
;      FILING DATE:  03-October-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
```

```
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4248.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2877 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join (126..203, 253..1776)
US-08-317-401E-1

  Query Match             7.1%;  Score 32.4;  DB 2;  Length 2877;
  Best Local Similarity   68.2%;  Pred. No. 1.1;
  Matches   45;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        275 TGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTA 334
              |||   ||||  |||| |   | || |||  ||  ||| ||||||| |||  |||| |
Db        383 TGAACAAGAAGCGTTTGATTCAGAAGGCCGTCTTCGATGAGCTGGTTTCCCTGGTTGATC 442

Qy        335 CCCATG 340
              ||||||
Db        443 CCCATG 448


Search completed: March 15, 2004, 10:15:12
Job time : 63.8 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 07:24:24 ; Search time 340.6 Seconds
(without alignments)
4928.974 Million cell updates/sec

Title: US-09-543-407-29
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	35.6	7.8	1395	15	US-10-375-932-214	Sequence 214, App
c	2	35.6	7.8	2025	15	US-10-375-932-268	Sequence 268, App
c	3	35.6	7.8	2031	15	US/10/375	Sequence 218, App
c	4	35.6	7.8	2031	15	US-10-375-932-266	Sequence 266, App
c	5	34.2	7.5	616	12	US-10-424-599-88945	Sequence 88945, A
c	6	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c	7	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c	8	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c	9	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c	10	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
c	11	33.6	7.4	388	9	US-09-960-352-8149	Sequence 8149, Ap
	12	33.6	7.4	3331	14	US-10-175-523-195	Sequence 195, App
	13	33.6	7.4	3331	14	US-10-177-293-189	Sequence 189, App
c	14	33.6	7.4	3778	15	US-10-369-493-27578	Sequence 27578, A
	15	33.6	7.4	4400	14	US-10-198-846-13928	Sequence 13928, A
	16	33.6	7.4	4635	14	US-10-198-846-11321	Sequence 11321, A
c	17	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
	18	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	19	33	7.2	407	9	US-09-960-352-4976	Sequence 4976, Ap
	20	32.6	7.1	129908	14	US-10-270-875-1	Sequence 1, Appli
	21	32.6	7.1	129908	14	US-10-270-878-1	Sequence 1, Appli
	22	32.6	7.1	129908	14	US-10-270-786-1	Sequence 1, Appli
	23	32.6	7.1	129908	14	US-10-270-710-1	Sequence 1, Appli
	24	32.6	7.1	129908	14	US-10-270-859-1	Sequence 1, Appli
	25	32.6	7.1	129908	14	US-10-270-846-1	Sequence 1, Appli
	26	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
	27	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c	28	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
c	29	32	7.0	1341	15	US-10-375-932-326	Sequence 326, App
c	30	32	7.0	1389	15	US-10-375-932-197	Sequence 197, App
	31	32	7.0	1584	12	US-10-425-114-22371	Sequence 22371, A
	32	31.8	7.0	827	9	US-09-956-004-36	Sequence 36, Appl
c	33	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
	34	31.8	7.0	1613	9	US-09-956-004-111	Sequence 111, App
	35	31.8	7.0	1772	14	US-10-234-432-3	Sequence 3, Appli
	36	31.8	7.0	1992	14	US-10-234-432-73	Sequence 73, Appl
	37	31.8	7.0	2151	14	US-10-234-432-5	Sequence 5, Appli
	38	31.8	7.0	2193	14	US-10-234-432-4	Sequence 4, Appli
	39	31.8	7.0	2243	14	US-10-234-432-2	Sequence 2, Appli
	40	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
	41	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
	42	31.6	6.9	1320	15	US-10-369-493-25834	Sequence 25834, A
	43	31.6	6.9	1482	15	US-10-369-493-25835	Sequence 25835, A
c	44	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
	45	31.4	6.9	339	15	US-10-316-194-84	Sequence 84, Appl

ALIGNMENTS

RESULT 1
US-10-375-932-214/c
; Sequence 214, Application US/10375932
; Publication No. US20040009469A1
; GENERAL INFORMATION:
; APPLICANT: Apt, Doris
; APPLICANT: Punnonen, Juha
; APPLICANT: Brinkman, Alice M.
; TITLE OF INVENTION: NOVEL FLAVIVIRUS ANTIGENS
; FILE REFERENCE: 0322.210US
; CURRENT APPLICATION NUMBER: US/10/375,932
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: US 60/360,030
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 345
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 214
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic polynucleotide: DEN-4 PRM15/tE CO parent
US-10-375-932-214

Query Match 7.8%; Score 35.6; DB 15; Length 1395;
Best Local Similarity 55.7%; Pred. No. 0.23;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
Qy      306 CTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAA 365
            | ||   || | |||| || ||| | || ||  || |   | || ||  |  ||| | |
Db      445 CGATCTGCACCAGGTTGCCGGTGATCTTGCCGCTGCAGCTGAACTTGGCGCAGGTCACCA 386

Qy      366 CGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTT 425
            |||||| || |   | ||| ||||| | | | ||   |    || |  || ||   |
Db      385 CGCCGCCCTTGCCGAACAGGCCGCAGCCGTTGCCCCAGCCCCTGTCCACCACGTCCCTCC 326

Qy      426 TG 427
            ||
```

```
Db        325 TG 324


RESULT 2
US-10-375-932-268/c
; Sequence 268, Application US/10375932
; Publication No. US20040009469A1
; GENERAL INFORMATION:
;  APPLICANT: Apt, Doris
;  APPLICANT:  Punnonen, Juha
;  APPLICANT:  Brinkman, Alice M.
;  TITLE OF INVENTION: NOVEL FLAVIVIRUS ANTIGENS
;  FILE REFERENCE: 0322.210US
;  CURRENT APPLICATION NUMBER: US/10/375,932
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: US 60/360,030
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 345
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 268
;   LENGTH: 2025
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic polynucleotide: 28E12( C15/full length prM/full length
US-10-375-932-268

  Query Match             7.8%;  Score 35.6;  DB 15;  Length 2025;
  Best Local Similarity   55.7%;  Pred. No. 0.28;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        313 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCG 372
              || |   || |||  | |||     ||| | | ||| |||||  |  ||| |  |||||||
Db        936 CACCACCTTGCCCTCGATGGGCTCCAGGCACTGGAACATGGCGCAGGTCACCACGCCGCC 877

Qy        373 CTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAAC 432
              || |   | ||| ||||| | | | ||   |     || |  || | ||   |||   ||
Db        876 CTTGCCGAACAGGCCGCAGCCGTTGCCCCAGCCCCTGTCCACCATGCTGTGCTTGCACAC 817

Qy        433 AA 434
              |
Db        816 GA 815


RESULT 3
US/10/375/c
; Sequence 218, Application US/10375932
; Publication No. US20040009469A1
; GENERAL INFORMATION:
;  APPLICANT: Apt, Doris
;  APPLICANT:  Punnonen, Juha
;  APPLICANT:  Brinkman, Alice M.
;  TITLE OF INVENTION: NOVEL FLAVIVIRUS ANTIGENS
;  FILE REFERENCE: 0322.210US
;  CURRENT APPLICATION NUMBER: US/10/375,932
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: US 60/360,030
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 345
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 218
;   LENGTH: 2031
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic polynucleotide: DEN-4 \C15/full prM/full E CO parent(a
on-of-the-nucleoti
US/10/375,932-218

  Query Match             7.8%;  Score 35.6;  DB 15;  Length 2031;
  Best Local Similarity   55.7%;  Pred. No. 0.28;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        306 CTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAA 365
              | ||   || | |||| || ||| | || ||  || |   | || ||  |  ||| | |
Db        943 CGATCTGCACCAGGTTGCCGGTGATCTTGCCGCTGCAGCTGAACTTGGCGCAGGTCACCA 884

Qy        366 CGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTT 425
              |||||| || |   | ||| ||||| | | | ||   |      || |  || ||   |
Db        883 CGCCGCCCTTGCCGAACAGGCCGCAGCCGTTGCCCCAGCCCCTGTCCACCACGTCCCTCC 824

Qy        426 TG 427
              ||
Db        823 TG 822


RESULT 4
US-10-375-932-266/c
; Sequence 266, Application US/10375932
; Publication No. US20040009469A1
; GENERAL INFORMATION:
;  APPLICANT: Apt, Doris
;  APPLICANT:  Punnonen, Juha
;  APPLICANT:  Brinkman, Alice M.
;  TITLE OF INVENTION: NOVEL FLAVIVIRUS ANTIGENS
;  FILE REFERENCE: 0322.210US
;  CURRENT APPLICATION NUMBER: US/10/375,932
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: US 60/360,030
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 345
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 266
;   LENGTH: 2031
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic polynucleotide: 28C1 (C15/full length prM/full length E
US-10-375-932-266

  Query Match             7.8%;  Score 35.6;  DB 15;  Length 2031;
  Best Local Similarity   55.7%;  Pred. No. 0.28;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        306 CTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAA 365
              | ||   || | |||| || ||| | || ||  || |   | || ||  |  ||| | |
Db        943 CGATCTGCACCAGGTTGCCGGTGATCTTGCCGCTGCAGCTGAACTTGGCGCAGGTCACCA 884

Qy        366 CGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTT 425
              |||||| || |   | ||| ||||| | | | ||   |      || |  || ||   |
Db        883 CGCCGCCCTTGCCGAACAGGCCGCAGCCGTTGCCCCAGCCCCTGTCCACCACGTCCCTCC 824

Qy        426 TG 427
              ||
Db        823 TG 822


RESULT 5
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
;   LENGTH: 616
;   TYPE: DNA
;   ORGANISM: Glycine max
```

```
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(616)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 616;
  Best Local Similarity   50.9%;  Pred. No. 0.45;
  Matches   81;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
              | | |  |||  ||  || ||| || || |||  | |    | |  | | | |   ||
Db        556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497

Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              | | ||   || | | |  || || | |   | | |||   ||  |  | |
Db        496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437

Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              || ||||||  |  |   ||| |||  ||||| ||| ||
Db        436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTA 398


RESULT 6
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Qiu, Dewen
;  APPLICANT:  Remick, Dean
;  TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
;  TITLE OF INVENTION:  RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
;  TITLE OF INVENTION:  DESICCATION
;  FILE REFERENCE: 21829/71
;  CURRENT APPLICATION NUMBER: US/09/835,684
;  CURRENT FILING DATE:  2001-04-16
;  PRIOR APPLICATION NUMBER: 60/198,359
;  PRIOR FILING DATE: 2000-04-19
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-835-684-6

  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.81;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 7
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  DeRocher, Jay
;  TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE: 21829/91
;  CURRENT APPLICATION NUMBER: US/09/880,371
;  CURRENT FILING DATE:  2001-06-13
;  PRIOR APPLICATION NUMBER: 60/211,585
;  PRIOR FILING DATE: 2000-06-15
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-880-371-6

  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.81;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 8
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
;  APPLICANT: Fan, Hao
;  APPLICANT:  Wei, Zhong-Min
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 21829/81
;  CURRENT APPLICATION NUMBER: US/09/879,248
;  CURRENT FILING DATE:  2001-06-12
;  PRIOR APPLICATION NUMBER: 60/212,211
;  PRIOR FILING DATE: 2000-06-16
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-879-248-5

  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.81;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 9
US-10-010-390-6/c
```

```
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Leon, Ernesto
;  APPLICANT:  Oviedo, Agustin
;  TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
;  TITLE OF INVENTION:  FROM ORNAMENTAL PLANTS
;  FILE REFERENCE: 21829/111
;  CURRENT APPLICATION NUMBER: US/10/010,390
;  CURRENT FILING DATE:  2001-11-05
;  PRIOR APPLICATION NUMBER: 60/248,169
;  PRIOR FILING DATE: 2000-11-13
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-010-390-6

  Query Match             7.5%;  Score 34;  DB 14;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.81;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
Db         89 GC 88


RESULT 10
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION:  RESISTANCE
;  FILE REFERENCE: 21829/203 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/10/441,736
;  CURRENT FILING DATE:  2003-05-20
;  PRIOR APPLICATION NUMBER: 60/107,243
;  PRIOR FILING DATE: 1998-11-05
;  PRIOR APPLICATION NUMBER: 09/431,614
;  PRIOR FILING DATE: 1999-11-02
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-441-736-5

  Query Match             7.5%;  Score 34;  DB 15;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.81;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
Db         89 GC 88


RESULT 11
US-09-960-352-8149/c
; Sequence 8149, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8149
;   LENGTH: 388
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 35-LIB3057-025-Q1-K1-A4
US-09-960-352-8149

  Query Match             7.4%;  Score 33.6;  DB 9;  Length 388;
  Best Local Similarity   49.4%;  Pred. No. 0.57;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        234 CGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAG 293
Db        297 CCCGGAGGTATGCCAGGCCCCGCAGAACCGCAATGCTGACCTTGCCCAGGATCTCTTCAG 238

Qy        294 AAATAATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 353
Db        237 GGATTCTCTTGGCTTCTTTCAACACCTGGTCCAGGGAGCCGCCATCCATGTGCTCCATGC 178

Qy        354 AGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGG 409
Db        177 AGATGCTGATCTCGCCGTCGCTGTAGAAGGCCCCGTAGAAGCCCACAATGTACGGG 122


RESULT 12
US-10-175-523-195
; Sequence 195, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
```

```
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 195
;   LENGTH: 3331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-175-523-195

  Query Match             7.4%;  Score 33.6;  DB 14;  Length 3331;
  Best Local Similarity   49.4%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||  | | || | || |  ||    | |  ||
Db       1277 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1336

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |   | | | ||   | | | | | | ||| |
Db       1337 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1396

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   | |||    || |  || | |    || ||
Db       1397 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1452


RESULT 13
US-10-177-293-189
; Sequence 189, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Gannavarpu, Manjula
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Mertens, Maureen
;  APPLICANT:  Myer, Vic
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Hoersch, Sebastian
;  APPLICANT:  Monahan, John
;  APPLICANT:  Meyers, Rachel E.
;  APPLICANT:  Bast Jr., Robert C.
;  APPLICANT:  Hortobagyi, Gabriel N.
;  APPLICANT:  Pusztai, Lajos
;  APPLICANT:  Meric, Funda
;  APPLICANT:  Sahin, Aysegul
;  APPLICANT:  Mills, Gordon B.
;  TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
;  TITLE OF INVENTION:  PREVENTION, AND THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-038
;  CURRENT APPLICATION NUMBER: US/10/177,293
;  CURRENT FILING DATE:  2002-06-21
;  PRIOR APPLICATION NUMBER: US 60/299,887
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/301,572
;  PRIOR FILING DATE: 2001-06-27
;  PRIOR APPLICATION NUMBER: US 60/306,501
;  PRIOR FILING DATE: 2001-07-18
;  PRIOR APPLICATION NUMBER: US 60/325,002
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/362,585
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/xxx,xxx
;  PRIOR FILING DATE: 2002-05-14
;  NUMBER OF SEQ ID NOS: 506
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 189
;   LENGTH: 3331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-177-293-189

  Query Match             7.4%;  Score 33.6;  DB 14;  Length 3331;
  Best Local Similarity   49.4%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||  | | || | || |  ||    | |  ||
Db       1277 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1336

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |   | | | ||   | | | | | | ||| |
Db       1337 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1396

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   | |||    || |  || | |    || ||
Db       1397 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1452


RESULT 14
US-10-369-493-27578/c
; Sequence 27578, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27578
;   LENGTH: 3778
;   TYPE: DNA
;   ORGANISM: Neurospora crassa
US-10-369-493-27578

  Query Match             7.4%;  Score 33.6;  DB 15;  Length 3778;
  Best Local Similarity   48.4%;  Pred. No. 1.9;
  Matches   93;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              |||||  |  | | |  |||||  |    ||    |   || ||       |   | | |
Db       1769 AAAGCATTCGCGGAAGTTCTGATTCTGTCACCGGCAGAAGGGGGCTGCCTGCGGTTGTCC 1710

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              | || | || |   | |||  | ||||  |  ||||||  || |  |||  |   | | |
Db       1709 AAGTCGACCTGCTGGGGGAACAGAGTATGAACGAACTGGTTCTGGCTGGCCTGGAAGAGA 1650

Qy        299 ATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCG 358
               |      |     | ||  |   | || |||| ||  || ||| |  |  ||  | |
Db       1649 TTCTGGACACCCTTGAGAAGCAAATCCTTGTTCTTGTCCGTGATACCGTCCACACTGTAC 1590

Qy        359 GTAATAACGCCG 370
              || | | | |||
Db       1589 GTCACATCACCG 1578


RESULT 15
US-10-198-846-13928
; Sequence 13928, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
```

```
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13928
;   LENGTH: 4400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-198-846-13928

  Query Match             7.4%;  Score 33.6;  DB 14;  Length 4400;
  Best Local Similarity   49.4%;  Pred. No. 2.1;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        259 AATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTG 318
              | ||  |  || || || |     |||||   | | || | || |  ||    | |  ||
Db       1440 ATTAACATCATGGAGCTCAAAACTAATGGGCCCCGGAAGATTGGCTACTGGAGTGAAGTG 1499

Qy        319 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCTGGTT 378
              |  |     |||||||||| |     |   | | | ||   | | | | | | ||| |
Db       1500 GACAAAATGGTTGTTACCCTTACTGAGCTCCCTTCTGGAAATGACACCTCTGGGCTTGAG 1559

Qy        379 AATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAA 434
              ||| |||| |    ||   |||   |  |||    || |  || | |    || ||
Db       1560 AATAAGACTGTTGTTGTCACCACAATTTTGGAATCTCCGTATGTTATGATGAAGAA 1615


Search completed: March 15, 2004, 14:16:42
Job time : 346.6 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 03:52:04 ; Search time 2500.2 Seconds
(without alignments)
6417.435 Million cell updates/sec

Title: US-09-543-407-29
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-29	Sequence 29, Appl
2	398.4	87.4	456	6	US-08-233-642A-56	Sequence 56, Appl
3	398.4	87.4	456	23	US-09-543-407-1	Sequence 1, Appli
4	377.6	82.8	456	23	US-09-543-407-23	Sequence 23, Appl
5	355.2	77.9	456	23	US-09-543-407-19	Sequence 19, Appl
6	348.8	76.5	456	23	US-09-543-407-13	Sequence 13, Appl
7	348.8	76.5	456	23	US-09-543-407-17	Sequence 17, Appl
8	348.8	76.5	456	23	US-09-543-407-21	Sequence 21, Appl
9	347.2	76.1	456	23	US-09-543-407-11	Sequence 11, Appl
10	347.2	76.1	456	23	US-09-543-407-25	Sequence 25, Appl
11	345.6	75.8	456	23	US-09-543-407-15	Sequence 15, Appl
12	344	75.4	456	23	US-09-543-407-27	Sequence 27, Appl
13	274.2	60.1	361	6	US-08-233-642A-54	Sequence 54, Appl
14	230.4	50.5	477	17	US-09-252-691-172	Sequence 172, App
15	230.4	50.5	477	17	US-09-252-691C-172	Sequence 172, App
16	230.4	50.5	477	51	US-10-417-886-172	Sequence 172, App
17	227.2	49.8	456	23	US-09-543-407-3	Sequence 3, Appli
18	227.2	49.8	675	13	US-08-978-878-3	Sequence 3, Appli
19	227.2	49.8	675	31	US-09-741-873B-3	Sequence 3, Appli
20	224.8	49.3	453	92	US-60-352-946-1	Sequence 1, Appli
21	224.8	49.3	453	101	US-60-444-371-1	Sequence 1, Appli
22	183.2	40.2	396	13	US-08-978-878-1	Sequence 1, Appli
23	158.4	34.7	360	31	US-09-741-873B-1	Sequence 1, Appli
24	55.4	12.1	522	17	US-09-252-691-170	Sequence 170, App
25	55.4	12.1	522	17	US-09-252-691C-170	Sequence 170, App
26	55.4	12.1	522	51	US-10-417-886-170	Sequence 170, App
c 27	49.4	10.8	78	23	US-09-543-407-52	Sequence 52, Appl
28	49	10.7	100	45	US-10-146-492B-77	Sequence 77, Appl
29	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
30	48	10.5	78	23	US-09-543-407-53	Sequence 53, Appl
31	42.6	9.3	78	23	US-09-543-407-49	Sequence 49, Appl
32	42.4	9.3	100	45	US-10-146-492B-79	Sequence 79, Appl
c 33	41	9.0	78	23	US-09-543-407-48	Sequence 48, Appl
c 34	38.8	8.5	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
35	37.4	8.2	1304	51	US-10-437-963-19773	Sequence 19773, A
36	36.4	8.0	25222	26	US-09-620-392-65403	Sequence 65403, A
37	35.8	7.9	561	17	US-09-252-691-171	Sequence 171, App
38	35.8	7.9	561	17	US-09-252-691C-171	Sequence 171, App
39	35.8	7.9	561	51	US-10-417-886-171	Sequence 171, App
c 40	35.6	7.8	1395	1	PCT-US03-05918-214	Sequence 214, App
c 41	35.6	7.8	1395	50	US-10-375-932-214	Sequence 214, App
c 42	35.6	7.8	2025	1	PCT-US03-05918-268	Sequence 268, App
c 43	35.6	7.8	2025	50	US-10-375-932-268	Sequence 268, App
c 44	35.6	7.8	2031	1	PCT/US03/05918-218	Sequence 218, App
c 45	35.6	7.8	2031	1	PCT-US03-05918-266	Sequence 266, App

ALIGNMENTS

```
RESULT 1
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 5.2e-135;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-642A-56

  Query Match             87.4%;  Score 398.4;  DB 6;  Length 456;
  Best Local Similarity   92.1%;  Pred. No. 1.6e-116;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |    |          |    | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             87.4%;  Score 398.4;  DB 23;  Length 456;
  Best Local Similarity   92.1%;  Pred. No. 1.6e-116;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |    |          |    | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             82.8%;  Score 377.6;  DB 23;  Length 456;
  Best Local Similarity   89.3%;  Pred. No. 7.5e-110;
  Matches  407;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |    |          |   ||     | | ||  |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 1.2e-102;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||| |||   |    ||||||    | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |    |                    |    | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 1.3e-100;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |    |                    |    | |  ||    |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 1.3e-100;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||      |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |    |                |  |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 1.3e-100;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |  |      | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||      || |||  |    |          |    | |  ||    ||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match            76.1%;  Score 347.2;  DB 23;  Length 456;
  Best Local Similarity  85.1%;  Pred. No. 4.3e-100;
  Matches  388;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||  || |||  |   |               |   | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||  |          | | || | ||   |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match            76.1%;  Score 347.2;  DB 23;  Length 456;
  Best Local Similarity  85.1%;  Pred. No. 4.3e-100;
  Matches  388;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |   |               |   | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match            75.8%;  Score 345.6;  DB 23;  Length 456;
  Best Local Similarity  84.9%;  Pred. No. 1.4e-99;
  Matches  387;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
                ||||||   || |||  |   |                |   | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
                ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

Query Match 75.4%; Score 344; DB 23; Length 456;
Best Local Similarity 84.6%; Pred. No. 4.6e-99;
Matches 386; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | |||  |     | |  ||   |    |      ||| ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300
```

```
Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || |||  |   |                |   | |  ||    ||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 60.1%; Score 274.2; DB 6; Length 361;
Best Local Similarity 88.7%; Pred. No. 1.1e-76;
Matches 297; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAAT 363
              |||   || |||  |   |                |   | |  ||    |||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 14
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match             50.5%;  Score 230.4;  DB 17;  Length 477;
  Best Local Similarity   69.1%;  Pred. No. 1.3e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| ||||||||||   | ||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||      |  |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  ||||||||| ||||||||   || |||| | |||||||||  ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | |||||| |||||||||||||   |||  ||||||||||||  |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || ||   |  || |   |      |  ||  |||    |||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||||  |||  | ||||| || || |  || |  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  |||||||  | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


RESULT 15
US-09-252-691C-172
; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-172

  Query Match             50.5%;  Score 230.4;  DB 17;  Length 477;
  Best Local Similarity   69.1%;  Pred. No. 1.3e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| ||||||||||   | ||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||      |  |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  ||||||||| ||||||||   || |||| | |||||||||  ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | |||||| |||||||||||||   |||  ||||||||||||  |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              ||||||   || ||   |  || |   |      |  ||  |||    |||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||||  |||  | ||||| || || |  || |  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  |||||||  | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


Search completed: March 15, 2004, 13:42:51
Job time : 2501.2 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 04:16:29 ; Search time 53.4 Seconds
(without alignments)
2945.701 Million cell updates/sec

Title: US-09-543-407-29
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 628836 seqs, 172478556 residues

Total number of hits satisfying chosen parameters: 1257672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	227.2	49.8	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	183.2	40.2	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33.8	7.4	150223	6	US-10-624-149A-1	Sequence 1, Appli
	4	30.6	6.7	398800	1	PCT-US03-32805-1	Sequence 1, Appli
c	5	29.6	6.5	201	7	US-60-548-091-10952	Sequence 10952, A
	6	29.6	6.5	878	6	US-10-100-683-11326	Sequence 11326, A
c	7	29.6	6.5	878	6	US-10-100-683-11986	Sequence 11986, A
	8	29.6	6.5	878	6	US-10-100-683-13006	Sequence 13006, A
	9	29.6	6.5	3990	5	US-09-830-230A-629	Sequence 629, App
c	10	29.6	6.5	79977	6	US-10-765-790-58	Sequence 58, Appl
c	11	29.6	6.5	187986	7	US-60-548-091-5629	Sequence 5629, Ap
	12	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
	13	29.4	6.4	1355	6	US-10-045-674A-593	Sequence 593, App
	14	29.2	6.4	1588	6	US-10-767-701-14522	Sequence 14522, A
	15	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	16	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
	17	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	18	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
	19	28.2	6.2	878	6	US-10-767-701-11365	Sequence 11365, A
c	20	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
	21	28.2	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	22	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
	23	28	6.1	420	6	US-10-767-795-3996	Sequence 3996, Ap
c	24	27.8	6.1	201	6	US-10-767-471-28241	Sequence 28241, A
	25	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
c	26	27.8	6.1	96539	6	US-10-767-471-10677	Sequence 10677, A
	27	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	28	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	29	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	30	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
	31	27.4	6.0	634	6	US-10-779-543-21897	Sequence 21897, A
c	32	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
	33	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
c	34	27.4	6.0	4398	6	US-10-775-169-113	Sequence 113, App
	35	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
	36	27.2	6.0	552	6	US-10-417-884A-3066	Sequence 3066, Ap
	37	27.2	6.0	752	6	US-10-767-795-3801	Sequence 3801, Ap
	38	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
	39	27.2	6.0	2034	6	US-10-781-014-229	Sequence 229, App
c	40	27.2	6.0	2880	6	US-10-100-683-5057	Sequence 5057, Ap
c	41	27.2	6.0	2881	6	US-10-100-683-2682	Sequence 2682, Ap
c	42	27.2	6.0	7380	6	US-10-417-375A-23	Sequence 23, Appl
c	43	27.2	6.0	7400	6	US-10-417-375A-19	Sequence 19, Appl
c	44	27.2	6.0	7460	6	US-10-417-375A-25	Sequence 25, Appl
c	45	27.2	6.0	113486	6	US-10-417-375A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 49.8%; Score 227.2; DB 5; Length 675;
Best Local Similarity 68.6%; Pred. No. 2.2e-67;
Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              || ||  |||||||  |      |      | |           |   |||    || |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873C
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873C-1

  Query Match             40.2%;  Score 183.2;  DB 5;  Length 396;
  Best Local Similarity   66.4%;  Pred. No. 1.4e-52;
  Matches  263;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||    | ||| |||||| |||||||||   || |||| |||| ||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
              || ||  |||||||  |      |      | |           |   |||    || |||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396


RESULT 3
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
;  APPLICANT: Neubauer, Antonie
;  APPLICANT:  Ziegler, Christina
;  TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
;  FILE REFERENCE: 1/1372
;  CURRENT APPLICATION NUMBER: US/10/624,149A
;  CURRENT FILING DATE:  2003-07-21
;  PRIOR APPLICATION NUMBER: 60/403,282
;  PRIOR FILING DATE: 2002-08-14
;  PRIOR APPLICATION NUMBER: DE 10233064
;  PRIOR FILING DATE: 2002-07-19
;  PRIOR APPLICATION NUMBER: DE 10317008
;  PRIOR FILING DATE: 2003-04-11
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
;   LENGTH: 150223
;   TYPE: DNA
;   ORGANISM: Equine herpesvirus 1
;  PUBLICATION INFORMATION:
;   AUTHORS: Telford,E.A.
;   AUTHORS: Watson, M.S.
;   AUTHORS: McBride, K.
;   AUTHORS: Davison, A.J.
;   TITLE: The DNA sequence of equine herpesvirus-1
;   JOURNAL: Virology
;   VOLUME: 189
;   ISSUE: 1
;   PAGES: 304-316
;   DATE: JUL-1992
;   DATABASE ACCESSION NUMBER: NC 001491, NCBI
;   DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

  Query Match             7.4%;  Score 33.8;  DB 6;  Length 150223;
  Best Local Similarity   51.7%;  Pred. No. 1.1;
  Matches   77;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              || ||  ||  |  | || ||  | ||| ||| |  |  ||| |||   | ||   |||
Db      95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy        233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
              |       |  |||| ||  ||||||    | || |  | | | ||| |    |  |||
Db      95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy        293 GAAATAATGCCACCTATGATCAGCTGGTT 321
              || |   |     | | || | | |||||
Db      95662 GAGAGCTTATTCACAACGAGCTGTTGGTT 95690


RESULT 4
PCT-US03-32805-1
; Sequence 1, Application PC/TUS0332805
; GENERAL INFORMATION:
```

```
; APPLICANT: deCODE genetics ehf.
; APPLICANT: Helgadottir, Anna
; APPLICANT: Gulcher, Jeffrey R.
; APPLICANT: Manolescu, Andrei
; TITLE OF INVENTION: Susceptibility Gene for Myocardial
; TITLE OF INVENTION: Infarction
; FILE REFERENCE: 2345.2048002
; CURRENT APPLICATION NUMBER: PCT/US03/32805
; CURRENT FILING DATE: 2003-10-16
; PRIOR APPLICATION NUMBER: 60/419,432
; PRIOR FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 535
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 398800
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (70677)...(70776)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (154988)...(155087)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 248521
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (269991)...(270091)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (272544)...(272643)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (279545)...(279644)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (300892)...(300991)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (327555)...(327654)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (332849)...(332948)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (341698)...(341698)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (345190)...(345289)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (350504)...(350603)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (391524)...(391623)
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (396625)...(396724)
; OTHER INFORMATION: n = A,T, C or G
```

```
PCT-US03-32805-1

  Query Match             6.7%;  Score 30.6;  DB 1;  Length 398800;
  Best Local Similarity   56.4%;  Pred. No. 20;
  Matches   57;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACC 324
              ||||||| ||||  |    ||| |  |||| | || ||| |  |  | |||| |  ||
Db      68535 ACTATTGCACTGTTTATAGATGTTGCCAGACAGAAAGCCCCAAAGCAGCAGCAGCATAGT 68594

Qy        325 CGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAA 365
                |  | || | || | ||| | ||  ||    || |||
Db      68595 ACCTTCCTAACACTTGGATTGGGAAATCTAGATTTTAGTAA 68635


RESULT 5
US-60-548-091-10952/c
; Sequence 10952, Application US/60548091
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STROKE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001506
; CURRENT APPLICATION NUMBER: US/60/548,091
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 24433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10952
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-548-091-10952

  Query Match             6.5%;  Score 29.6;  DB 7;  Length 201;
  Best Local Similarity   53.6%;  Pred. No. 1.3;
  Matches   59;  Conservative    1;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGC 316
              | || || |  | | ||   |  |     || | ||  ||| |   ||||  | |  ||||
Db        190 AAAAGAGAAGAACTTAATGAATGCTTTCTGCTCTCTCAAAGATCACCACACACGGACAGC 131

Qy        317 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAAC 366
                | |   |||  || | | ||| || | :  |  |||||||||  |  ||
Db        130 CAGGTCAGCGTCCTGCTCCTCATCAATTCRGGCGGGCAGGCGGTGAAGAC 81


RESULT 6
US-10-100-683-11326
; Sequence 11326, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
```

```
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11326
;   LENGTH: 878
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (741)..(741)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-100-683-11326

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 878;
  Best Local Similarity   56.0%;  Pred. No. 2.6;
  Matches   56;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCG 326
              | |||  |    |  | || | | | | || ||| |    |  | ||||||| ||| | | ||
Db        627 TTTTGCTCACCCAGAAAACGCTGTGAAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACG 686

Qy        327 TGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAAC 366
              || | |  |   ||| |||  |       ||||||| |||
Db        687 AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAAAAC 726


RESULT 7
US-10-100-683-11986/c
; Sequence 11986, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11986
;   LENGTH: 878
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (138)..(138)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-100-683-11986

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 878;
  Best Local Similarity   56.0%;  Pred. No. 2.6;
  Matches   56;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCG 326
              | |||  |    |  | || | | | | || ||| |    |  | ||||||| ||| | | ||
Db        252 TTTTGCTCACCCAGAAAACGCTGTGAAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACG 193

Qy        327 TGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAAC 366
              || | |  |   ||| |||  |       ||||||| |||
Db        192 AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAAAAC 153


RESULT 8
US-10-100-683-13006
; Sequence 13006, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13006
;   LENGTH: 878
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (741)..(741)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-100-683-13006

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 878;
  Best Local Similarity   56.0%;  Pred. No. 2.6;
  Matches   56;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCG 326
              | |||  |    |  | || | | | | || ||| |    |  | ||||||| ||| | | ||
Db        627 TTTTGCTCACCCAGAAAACGCTGTGAAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACG 686

Qy        327 TGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAAC 366
              || | |  |   ||| |||  |       ||||||| |||
Db        687 AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAAAAC 726


RESULT 9
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Lyme Disease Vaccines
;  FILE REFERENCE: PB481US
;  CURRENT APPLICATION NUMBER: US/09/830,230A
```

```
;  CURRENT FILING DATE:  2001-09-27
;  PRIOR APPLICATION NUMBER: PCT/US98/12718
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/057,483
;  PRIOR FILING DATE: 1997-09-03
;  PRIOR APPLICATION NUMBER: 60/053,344
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/053,377
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/050,359
;  PRIOR FILING DATE: 1997-06-20
;  NUMBER OF SEQ ID NOS: 756
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
;   LENGTH: 3990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1139)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1143)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1148)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1210)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1244)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1247)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1250)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1251)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

  Query Match             6.5%;  Score 29.6;  DB 5;  Length 3990;
  Best Local Similarity   51.5%;  Pred. No. 5.3;
  Matches   68;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||      ||  ||| |  | | |  |  | ||
Db       3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |   |  | ||||||||| | |||    |||  ||
Db       3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy        277 ACTCAGAATGGT 288
              |   || ||| |
Db       3553 AAGAAGGATGAT 3564


RESULT 10
US-10-765-790-58/c
; Sequence 58, Application US/10765790
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
;   LENGTH: 79977
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-58

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 79977;
  Best Local Similarity   50.0%;  Pred. No. 21;
  Matches   74;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              ||| ||| | | || | | ||  ||  | ||||  ||| ||  |  |  |||       |
Db      15123 GGTGATGATGAAGGAGATGGTGATGGTGATGGTGGTGATGATGATGATGGTGATGATGGT 15064

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 339
              || |||||    |  || |||       ||| | |    | |    ||| || |   |||
Db      15063 GAGGATGGTGATGGTGATGATGGAGGAGATGGTGACGGCGATGGTGGTGATGGTGATCAT 15004

Qy        340 GAAATGGCACATGCAGGCGGTAATAACG 367
              ||    |   |||  |  |||  | | |
Db      15003 GATGGTGATGATGGTGATGGTGGTGAGG 14976


RESULT 11
US-60-548-091-5629/c
; Sequence 5629, Application US/60548091
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STROKE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001506
;  CURRENT APPLICATION NUMBER: US/60/548,091
;  CURRENT FILING DATE:  2004-02-27
;  NUMBER OF SEQ ID NOS: 24433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5629
;   LENGTH: 187986
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-548-091-5629

  Query Match             6.5%;  Score 29.6;  DB 7;  Length 187986;
  Best Local Similarity   53.6%;  Pred. No. 31;
  Matches   59;  Conservative    1;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCAGC 316
              | || || |  | | ||   |  |    || | ||  ||| |   ||||  | |  ||||
Db      55805 AAAAGAGAAGAACTTAATGAATGCTTTCTGCTCTCTCAAAGATCACCACACACGGACAGC 55746

Qy        317 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAAC 366
              | |   |||  || | | ||| || | :  |  ||||||||| |  ||
Db      55745 CAGGTCAGCGTCCTGCTCCTCATCAATTCRGGCGGGCAGGCGGTGAAGAC 55696


RESULT 12
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
```

```
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 1.9;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||    |       ||||| | |     | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
              |||||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 13
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674A
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 06/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674A-593

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 1355;
  Best Local Similarity   25.2%;  Pred. No. 3.8;
  Matches   39;  Conservative   50;  Mismatches   66;  Indels    0;  Gaps    0;

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137
              :||:||:||: ::  :  :  : ::||:||:  ::::   ||:   | :::   :    |
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197
              :  :: :  :  :||::::| :: :| :|   |:| :    |:|     ||   :
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              : |:  : |:|| ||: |:| :| : |:|  ||:|
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYG 922


RESULT 14
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
;   LENGTH: 1588
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 1588;
  Best Local Similarity   44.4%;  Pred. No. 4.8;
  Matches  118;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
              | |  |||| |       | ||| | |  |       | ||||| |||  |  | |  |
Db        455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
              |||||  ||| || ||   |||   |     |  |   |||          ||||  |  |
Db        515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy        160 AACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
                | ||   ||| |    |  | | |||  ||  | |||  |   | | || |     |
Db        575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              |||  |   | |  |||         |  |||||  || | | ||  |  || |    ||
Db        635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy        280 CAGAATGGTTTCAGAAATAATGCCAC 305
                | | ||  |   |  | ||||
Db        695 GCTGGTAGCTTTGGCGGTGGTTCCAC 720


RESULT 15
PCT-US04-05654-1030
; Sequence 1030, Application PC/TUS0405654
; GENERAL INFORMATION:
;  APPLICANT: Sherman, Bradley K
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline  E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Dubell III, Arnold N
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Repetti, Peter
;  APPLICANT:  Century, Karen
;  APPLICANT:  Gutterson, Neal
```

; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E
; APPLICANT: Kumimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1030
; LENGTH: 1186
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
PCT-US04-05654-1030

Query Match 6.3%; Score 28.6; DB 1; Length 1186;
Best Local Similarity 59.0%; Pred. No. 6.6;
Matches 49; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy         44 CTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCG 103
              | || | |||||  || | || ||      |  | || ||||||||   |  |   ||||
Db        606 CAGGGACTGCTCCAGCGGCCGGCGGCGGCTACGGCGGTGGCGGCGGCGGTGGTGGTGGCG 665

Qy        104 GCGGCAATAGTTCCGGCCCGGAC 126
              || || | |  | ||||| || |
Db        666 GCTGCTACAACTGCGGCCAGGCC 688
```

Search completed: March 15, 2004, 13:47:30
Job time : 55.4 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 02:55:53 ; Search time 1963.4 Seconds
(without alignments)
6935.497 Million cell updates/sec

Title: US-09-543-407-29
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	227.8	50.0	680	12	BJ618688	BJ618688 BJ618688
	2	37.6	8.2	509	14	CF368512	CF368512 853167 MA
	3	37.6	8.2	601	28	AZ460219	AZ460219 1M0265A02
c	4	37.2	8.2	968	14	CF596693	CF596693 AGENCOURT
c	5	34.8	7.6	764	14	CA752110	CA752110 UI-M-FD0-
	6	34.6	7.6	458	9	AU193911	AU193911 AU193911
	7	34.6	7.6	499	9	AV429805	AV429805 AV429805
	8	34.6	7.6	505	9	AU190596	AU190596 AU190596
	9	34.6	7.6	517	10	BF429785	BF429785 255670 MA
	10	34.6	7.6	530	9	AU196255	AU196255 AU196255
c	11	34.2	7.5	413	28	AQ186498	AQ186498 HS_3090_B
c	12	34.2	7.5	971	13	BQ891514	BQ891514 AGENCOURT
	13	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	14	34	7.5	234	10	BB581987	BB581987 BB581987
c	15	34	7.5	324	9	AA083436	AA083436 zn09a08.r
c	16	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
	17	34	7.5	656	14	CB557442	CB557442 AMGNNUC:C
	18	34	7.5	1530	14	CF111010	CF111010 Shultzomi
c	19	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
	20	33.8	7.4	463	13	BX551478	BX551478 BX551478
c	21	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	22	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
c	23	33.8	7.4	664	28	BH988171	BH988171 oeh29g06.
	24	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	25	33.8	7.4	1097	28	BH770962	BH770962 LLMGtag69
c	26	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	27	33.8	7.4	1228	29	CG753535	CG753535 P048-3-F0
c	28	33.8	7.4	2313	11	AK028764	AK028764 Mus muscu
c	29	33.6	7.4	385	9	AA983039	AA983039 vx58b05.r
c	30	33.6	7.4	386	10	AW647538	AW647538 EST326081
	31	33.6	7.4	393	10	AW464937	AW464937 BP230017A
c	32	33.6	7.4	423	10	AW325822	AW325822 16823 MAR
c	33	33.6	7.4	552	10	BF193938	BF193938 245679 MA
c	34	33.6	7.4	603	12	BI683390	BI683390 464752 MA
c	35	33.6	7.4	661	28	CC096651	CC096651 CSU-K34.1
c	36	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	37	33.6	7.4	791	13	BU057696	BU057696 UI-M-FR0-
c	38	33.4	7.3	505	6	AL809250	Al809250 Triticum
c	39	33.4	7.3	699	12	BI959328	BI959328 HVSMEn001
	40	33.4	7.3	715	13	BU444683	BU444683 603765715
	41	33.4	7.3	751	28	BH490396	BH490396 BOGYZ57TF
	42	33.2	7.3	629	9	AU295325	AU295325 AU295325
	43	33.2	7.3	788	9	AU139321	AU139321 AU139321
	44	33.2	7.3	977	29	CNS034F7	AL227356 Tetraodon
c	45	33.2	7.3	1646	10	BF579120	BF579120 602096054

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 50.0%; Score 227.8; DB 12; Length 680;
Best Local Similarity 68.6%; Pred. No. 6.1e-54;
Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db 533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db 413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db 353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy 301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360
Db 293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy 361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db 233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy 421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db 173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138

RESULT 2
CF368512
LOCUS CF368512 509 bp mRNA linear EST 25-AUG-2003
DEFINITION 853167 MARC 3PIG Sus scrofa cDNA 3', mRNA sequence.
ACCESSION CF368512
VERSION CF368512.1 GI:34174101
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 509)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J., Wise,T.A., Nonneman,D.J., Wray,J.E. and Keele,J.W.
TITLE A second set of porcine ESTs from a pooled-tissue normalized library
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329.
Plate: SRG8025 row: F column: 11
Seq primer: TAGAAGGCACAGTCGAGG.
FEATURES Location/Qualifiers
source 1. .509
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3PIG"
/note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI; Library made with RNA pooled from multiple tissues including brain, liver, muscle, placenta/endometrium, ovary, testes, and bone marrow."
ORIGIN

Query Match 8.2%; Score 37.6; DB 14; Length 509;
Best Local Similarity 52.6%; Pred. No. 9.5;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 286 GGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 345
Db 98 GGTTCATGAAAGGACGCCTGCTTTCCACAGGGGCCTGCCTTTTCAGGCCCCTAGGAGATG 157

Qy 346 GCACATGCAGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTA 405
Db 158 GAACAGGATGGTGGTTAAGAGCCCAGGCTCTGCAGTCAGAGTGACTGGGATGGGATCCTG 217

Qy 406 ATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACG 441
Db 218 GCTCTACTTAAGGTTAGCTTTGTGACCCTAGGCAAG 253

RESULT 3
AZ460219
LOCUS AZ460219 601 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0265A02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0265A02 R, genomic survey sequence.
ACCESSION AZ460219
VERSION AZ460219.1 GI:10618344
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0265 row: A column: 02
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 601.
FEATURES Location/Qualifiers
source 1. .601
/organism="Mus musculus"
/mol_type="genomic DNA"

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0265A02"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 8.2%; Score 37.6; DB 28; Length 601;
Best Local Similarity 55.3%; Pred. No. 10;
Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy        247 CAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACC 306
              ||| ||||  |||  |  |  ||  |    ||| | ||   | | ||||||  |   |
Db        204 CAGTGTGCTCATATAAACAAAATAAATAAAACTTAAAAAAATATTAGAAATCACTTCTGG 263

Qy        307 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAAC 366
                | ||  ||||  | || ||| |   | |  |   | |||| | ||||| ||  ||||
Db        264 ATTAATTGGCTGCATGCCTGTGATCCCAGCACTTGCAAGGCAGAGGCAGGTGGATATAAG 323

Qy        367 GCCGCGCTGGTT 378
              |||   || |||
Db        324 GCCAGCCTTGTT 335
```

RESULT 4
CF596693/c
LOCUS CF596693 968 bp mRNA linear EST 26-SEP-2003
DEFINITION AGENCOURT_15668817 NICHD_Hs_Ov1 Homo sapiens cDNA clone IMAGE:30704437 5', mRNA sequence.
ACCESSION CF596693
VERSION CF596693.1 GI:36353386
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 968)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gregory F. Erickson, Ph.D.
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM253 row: h column: 14
High quality sequence stop: 411.
FEATURES Location/Qualifiers
source 1. .968
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30704437"
/tissue_type="Ovary"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NICHD_Hs_Ov1"
/note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); Library is oligo-dT primed and directionally cloned. Granulosa lutein cells aspirated from preovulatory folicles of normal cycling women undergoing ovulation induction for infertility due to male factor and normal doners. The cells were from follicles stimulated with Lupron, FSH and hCG. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 2.23 kb (range 1.0-4.5 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match 8.2%; Score 37.2; DB 14; Length 968;
Best Local Similarity 53.4%; Pred. No. 16;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy        186 TGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGG 245
              | ||   ||| | | || | |    ||||||    |  | ||   | | ||| |   ||
Db        592 TTCCACAAAAGCGGGAAAGTCATAAACCCAGGAAAGACAAGGAGGCCGGGCCAACCAAGC 533

Qy        246 CCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCAC 305
              ||  | |  || |  |   | |||  |||  | |    | | |   |||    ||  || |
Db        532 CCTAGTTCTGGGTCCTTAGAATATGCAACAAAATGGACAAGTTGGGAGATGGAACCCCCC 473

Qy        306 CTATGATCAGCTGGTTACCCGTGTTG 331
              |||| ||||||| ||| |||  | ||
Db        472 CTATCATCAGCTAGTTTCCCAAGGTG 447
```

RESULT 5
CA752110/c
LOCUS CA752110 764 bp mRNA linear EST 09-JUL-2003
DEFINITION UI-M-FD0-cdl-b-16-0-UI.r1 NIH_BMAP_FD0 Mus musculus cDNA clone IMAGE:6832145 5', mRNA sequence.
ACCESSION CA752110
VERSION CA752110.1 GI:25582043
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 764)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, Univeristy of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
The following repetitive elements were found in this cDNA sequence: 26-59, >GC_rich#Low_complexity
Seq primer: pYX-5.

```
FEATURES             Location/Qualifiers
     source          1. .764
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6832145"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FD0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA
                     tail, is TGAGAGAGCC. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 14;  Length 764;
  Best Local Similarity   52.0%;  Pred. No. 71;
  Matches   78;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTC 280
              | | ||||     | | || ||||||||||    |||| | | | | ||| || | |
Db        671 GATCTGGTCTTCTCTCGGAGGTAGGCCAGGCCCCGGAGCACCGCAATGCTGACCTTCCCC 612

Qy        281 AGAATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATG 340
              |  |||  ||||| |||         |     ||||    |   ||     |  |||
Db        611 AAGATGTCTTCAGGAATCCGCTTGGCCTCCTTCAGTACCTGGTCCAGTGAGCCACCATCC 552

Qy        341 AAATGGCACATGCAGGCGGTAATAACGCCG 370
              |  ||   |||||||  | | ||  |||||
Db        551 ATGTGCTCCATGCAGATGCTGATCTCGCCG 522


RESULT 6
AU193911
LOCUS       AU193911                458 bp    mRNA    linear   EST 14-OCT-2003
DEFINITION  AU193911 Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis
            cDNA clone PFL052f02_r 5', mRNA sequence.
ACCESSION   AU193911
VERSION     AU193911.1  GI:31934023
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and
            Tabata,S.
  TITLE     COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS
            OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG
            FREQUENCY ANALYSIS
  JOURNAL   J. Phycol. 39 (5), 923-930 (2003)
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Porphyra yezoensis"
                     /mol_type="mRNA"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PFL052f02_r"
                     /dev_stage="sporophytes"
                     /clone_lib="Porphyra yezoensis TU-1 sporophytes"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 9;  Length 458;
  Best Local Similarity   49.2%;  Pred. No. 65;
  Matches   91;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        195 ATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGC 254
              |||||    |  | |   ||||  ||  | ||   | | ||  || | ||   |||||||
Db        270 ATCTGCTCTGGACGTCTTCCAGCACGTCTTTGACCAGGACGATGACGAAGCGGAGGGTGC 329

Qy        255 GGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCA 314
              | | |||| || | | | ||           | |  |     |  ||| | |||||||  |
Db        330 CGGTGATAGGACCACTAAGCTATGGTCCCGCGCTCCCGCGGACGATGGCGCCTATGACGA 389

Qy        315 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCT 374
              | ||  |   ||   |   ||  |||   || | ||  | |  ||||||  |
Db        390 GACGGCGGCGGTCGTGTCAAGTGGCGATGCGGCGGTGGCGGACGAAATTCTCGCCGCCAT 449

Qy        375 GGTTA 379
              |  ||
Db        450 GCCTA 454


RESULT 7
AV429805
LOCUS       AV429805                499 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV429805 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PL008g07_r 5', mRNA sequence.
ACCESSION   AV429805
VERSION     AV429805.1  GI:8585030
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
   PUBMED   10907854
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Porphyra yezoensis"
                     /mol_type="mRNA"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PL008g07_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 9;  Length 499;
  Best Local Similarity   49.2%;  Pred. No. 68;
  Matches   91;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        195 ATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGC 254
              |||||    |  | |   ||||  ||  | ||   | | ||  || | ||   |||||||
Db        299 ATCTGCTCTGGACGTCTTCCAGCACGTCTTTGACCAGGACGATGACGAAGCGGAGGGTGC 358
```

```
Qy        255 GGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCA 314
              | | |||| || | | | ||           | |  |    |  ||| | |||||||  |
Db        359 CGGTGATAGGACCACTAAGCTATGGTCCCGCGCTCCCGCGGACGATGGCGCCTATGACAA 418

Qy        315 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCT 374
              |  ||   |    ||    |      ||   |||    || | ||  | |  ||||||  |
Db        419 GACGGCGGCGGTCGTGTCAAGTGGCGATGCGGCGGTGGCGGACGAAATTCTCGCCGCCAT 478

Qy        375 GGTTA 379
              |  ||
Db        479 GCCTA 483


RESULT 8
AU190596
LOCUS       AU190596                505 bp    mRNA    linear   EST 14-OCT-2003
DEFINITION  AU190596 Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis
            cDNA clone PFL004f08_r 5', mRNA sequence.
ACCESSION   AU190596
VERSION     AU190596.1  GI:31927398
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and
            Tabata,S.
  TITLE     COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS
            OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG
            FREQUENCY ANALYSIS
  JOURNAL   J. Phycol. 39 (5), 923-930 (2003)
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Porphyra yezoensis"
                     /mol_type="mRNA"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PFL004f08_r"
                     /dev_stage="sporophytes"
                     /clone_lib="Porphyra yezoensis TU-1 sporophytes"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 9;  Length 505;
  Best Local Similarity   49.2%;  Pred. No. 68;
  Matches   91;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        195 ATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGC 254
              |||||    |  | |   |||| ||  | ||   | | ||  || | ||   |||||||
Db        295 ATCTGCTCTGGACGTCTTCCAGCACGTCTTTGACCAGGACGATGACGAAGCGGAGGGTGC 354

Qy        255 GGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCA 314
              | | |||| || | | | ||           | |  |    |  ||| | |||||||  |
Db        355 CGGTGATAGGACCACTAAGCTATGGTCCCGCGCTCCCGCGGACGATGGCGCCTATGACGA 414

Qy        315 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCT 374
              |  ||   |    ||    |      ||   |||    || | ||  | |  ||||||  |
Db        415 GACGGCGGCGGTCGTGTCAAGTGGCGATGCGGCGGTGGCGGACGAAATTCTCGCCGCCAT 474

Qy        375 GGTTA 379
              |  ||
Db        475 GCCTA 479


RESULT 9
BF429785
LOCUS       BF429785                517 bp    mRNA    linear   EST 29-NOV-2000
DEFINITION  255670 MARC BSM Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF429785
VERSION     BF429785.1  GI:11441882
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Warren,W.C., Tao,N., Allison,T., Wagner,S., Mathialagan,N.,
            Kata,S., Johnson,J., Smith,T.P.L. and Womack,J.
  TITLE     A survey of genes transcribed in bovine skeletal muscle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: GGAAACAGCTATGACCATG
            BACKWARD: GTTTTCCCAGTCACGAC
            Seq primer: AATTAACCCTCACTAAAGGG.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="Skeletal muscle"
                     /lab_host="XLOLR"
                     /clone_lib="MARC BSM"
                     /note="Vector: Uni-ZAP XR; Site_1: EcoRI; Site_2: XhoI;
                     Library obtained from Stratagene, catalog #937721. Library
                     made from skeletal muscle of a two year old Holstein cow."
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 10;  Length 517;
  Best Local Similarity   61.8%;  Pred. No. 69;
  Matches   55;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy        283 AATGGTTTCAGAAATAATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAA 342
              | |||  |  |||||| ||  ||  ||  ||| | | || ||    || |   |||||||
Db         32 ACTGGACTTTGAAATAGTGTGACTGATATTCATCAGCTTTCCTTCCTTCTGGGCCATGAA 91

Qy        343 ATGGCACATGCAGGCGGTAATAACGCCGC 371
              || ||||||||||    | |  | || ||
Db         92 ATAGCACATGCAGTACTTGAGCATGCTGC 120


RESULT 10
AU196255
LOCUS       AU196255                530 bp    mRNA    linear   EST 15-OCT-2003
DEFINITION  AU196255 Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis
            cDNA clone PFL087b07_r 5', mRNA sequence.
ACCESSION   AU196255
VERSION     AU196255.1  GI:31938711
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and
            Tabata,S.
  TITLE     COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS
            OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG
            FREQUENCY ANALYSIS
  JOURNAL   J. Phycol. 39 (5), 923-930 (2003)
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
```

```
           Kazusa DNA Research Institute
           Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
           Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Porphyra yezoensis"
                     /mol_type="mRNA"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PFL087b07_r"
                     /dev_stage="sporophytes"
                     /clone_lib="Porphyra yezoensis TU-1 sporophytes"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 9;  Length 530;
  Best Local Similarity   49.2%;  Pred. No. 69;
  Matches   91;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        195 ATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGC 254
              |||||     |  | |    ||||  ||  | ||   | | ||  || | ||   |||||||
Db        155 ATCTGCTCTGGACGTCTTCCAGCACGTCTTTGACCAGGACGATGACGAAGCGGAGGGTGC 214

Qy        255 GGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATGATCA 314
              | | |||| || | | | ||          | |  |     |  ||| | |||||||  |
Db        215 CGGTGATAGGACCACTAAGCTATGGTCCCGCGCTCCCGCGGACGATGGCGCCTATGACGA 274

Qy        315 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCGCGCT 374
              |  ||   |    ||    |     ||  |||   || | ||  | |  ||||||  |
Db        275 GACGGCGGCGGTCGTGTCAAGTGGCGATGCGGCGGTGGCGGACGAAATTCTCGCCGCCAT 334

Qy        375 GGTTA 379
              |  ||
Db        335 GCCTA 339


RESULT 11
AQ186498/c
LOCUS       AQ186498                413 bp    DNA     linear   GSS 01-NOV-1998
DEFINITION  HS_3090_B1_C05_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3090 Col=9 Row=F, genomic survey
            sequence.
ACCESSION   AQ186498
VERSION     AQ186498.1  GI:3585940
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3090  row: F  column: 9
            Class: BAC ends
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3090 Col=9 Row=F"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 28;  Length 413;
  Best Local Similarity   51.3%;  Pred. No. 81;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCTATG 310
              |||  ||| ||||||  |  |    || ||    | ||  || || | |||| |      |||
Db        199 GTGGAGATGATAGTAATGCTTTTCTCACGTGGCATTATTGCAAAGATAAGGTTTATTATA 140

Qy        311 ATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTAATAACGCCG 370
              | ||  | ||||      |   ||    |  || |  ||||||  ||    ||||||
Db        139 AGCATATCGTTATTGTAGAGATTCAGCGGGAGAGTGCACATCAAGCACTTAATAAGTAGC 80

Qy        371 CGCTGGTTAATCAGACCGCATCTGATTCCAGC 402
                ||  || | |    | ||    |||  |||
Db         79 TAGTGTATAGTAAATACNCAGTGCATTGTAGC 48


RESULT 12
BQ891514/c
LOCUS       BQ891514                971 bp    mRNA    linear   EST 16-AUG-2002
DEFINITION  AGENCOURT_8747984 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335378
            5', mRNA sequence.
ACCESSION   BQ891514
VERSION     BQ891514.1  GI:22283528
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 971)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13797  row: l  column: 03
            High quality sequence stop: 683.
FEATURES             Location/Qualifiers
     source          1. .971
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6335378"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_130"
                     /note="Organ: otocysts; Vector: pCMV-SPORT6.1; Site_1:
                     EcoRV; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  Average insert size 1.95  kb.  Constructed by
                     ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 13;  Length 971;
  Best Local Similarity   52.6%;  Pred. No. 1.2e+02;
  Matches   72;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        234 CGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAG 293
              | | || ||||||||||    |||| |  | | |  ||| ||  | ||  |||  |||||
Db        757 CTCGGAGGTAGGCCAGGCNCCGGAGCACCGCAATGCTGACCTCNCCCAAGATGTCTTCAG 698
```

```
Qy        294 AAATAATGCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 353
              |||         |    ||||    |   ||     |  |||    |  ||   |||||
Db        697 GAATCCGCTTGGCCTCCTTCAGTACCTGGTCCAGTGAGCCACCATCCATGTGCTCCATGC 638

Qy        354 AGGCGGTAATAACGCCG 370
              ||  | | ||  |||||
Db        637 AGATGCTGATCTCGCCG 621
```

RESULT 13
BG671547

```
LOCUS       BG671547               1036 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5',
            mRNA sequence.
ACCESSION   BG671547
VERSION     BG671547.1  GI:13893646
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C.,
            Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D.,
            Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
   PUBMED   12060780
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1036
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNBUB05"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /clone_lib="Rat DRG Library"
ORIGIN
```

```
  Query Match             7.5%;  Score 34.2;  DB 12;  Length 1036;
  Best Local Similarity   49.2%;  Pred. No. 1.2e+02;
  Matches   90;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy         11 TAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTC 70
              | ||||   ||  | |  |||    |||  || | | || |||| |     |||||  |
Db        168 TGAAAGATACATGAGTGACAGTGTCCGTGATTGCAGACATTGCTGTATGATGCGTCCTGT 227

Qy         71 CACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAA 130
              |   | | | |  ||  |||||  ||  ||   | |   || | |   |   ||     |
Db        228 CTGTAGGAGTCCTCGCTGGTAACGATGCCGAGAGTGCACATCGATATCGGTGGGCGGTTA 287

Qy        131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
              | |       |  |  ||   | |     | ||| |||   ||| ||||  ||| || ||
Db        288 CATCTGAAGTGGAAGAGGTGGCTCTTTCGAACTCTGACTTTGATCTGTCAAGGGACGCTT 347

Qy        191 GTA 193
              |||
Db        348 GTA 350
```

RESULT 14
BB581987/c

```
LOCUS       BB581987               234 bp    mRNA    linear   EST 30-NOV-2000
DEFINITION  BB581987 RIKEN full-length enriched, adult male colon Mus musculus
            cDNA clone 9030005O05 5', mRNA sequence.
ACCESSION   BB581987
VERSION     BB581987.1  GI:11478531
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
            Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
            Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
            Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
            Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
            Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9030005O05"
                     /sex="male"
                     /tissue_type="colon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male colon"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
```

Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match 7.5%; Score 34; DB 10; Length 234;
Best Local Similarity 51.3%; Pred. No. 73;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy         35 TCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATC 94
              | ||  ||  | |    | |   || ||  || |    ||   || | || || ||    |
Db        200 TAGTTTTTGGTAGGTCGGTTTATGCGGGTTTCTTCTTCCACCAGGCCCGCCGCCGTTTCC 141

Qy         95 ATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTT 154
                ||||  |||  ||  ||  ||       |  || | ||||||| ||  ||  ||
Db        140 CGAACGCCGGGTGCTTTATAGCCTTGTTCAAAATATGGTAGCGCATTTACCACCACCTTT 81

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              |||  | ||   || | |||  |  ||  |||  ||
Db         80 CCGGTTACCAAGCCCGTGCCCGCCATTGCGCGGC 47
```

RESULT 15
AA083436/c
LOCUS AA083436 324 bp mRNA linear EST 23-DEC-1997
DEFINITION zn09a08.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone IMAGE:546902 5' similar to TR:G988221 G988221 TBC1. ;, mRNA sequence.
ACCESSION AA083436
VERSION AA083436.1 GI:1625497
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 324)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham.

FEATURES Location/Qualifiers
source 1. .324
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3925092"
/db_xref="taxon:9606"
/clone="IMAGE:546902"
/dev_stage="hNT neurons"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene hNT neuron (#937233)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Differentiated, post mitotic hNT neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 7.5%; Score 34; DB 9; Length 324;
Best Local Similarity 50.6%; Pred. No. 83;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        242 TAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
              | | | |  |||  ||  ||||| | |||   ||    |    |||  |    |||| ||
Db        236 TGGCCAATCGTGAAGAGCATAGTTCGATTTTCCTCGATATCTGTGGGCTGCACAATATTG 177

Qy        302 CCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGTA 361
                ||  | ||  || |||||  |  ||| | |  ||  |||   ||  | | || || |
Db        176 TGTCCGCTAATGAGGTGGTTCTCAATGTCGCTCTCCTCGAAGGAGCTGAAGAAGCCGCTG 117

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCG 403
              |   |  || |   |    |||  |   || |  ||||||
Db        116 CTTCGGAGGCCCCCATCCTGCAGCTCCTTCCTAAAGCCAGCG 75
```

Search completed: March 15, 2004, 10:09:35
Job time : 1967.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		456	100.0	456	6	AX814811	AX814811 Sequence
2		456	100.0	648	1	ECOCSGAA	L04979 Escherichia
3		454.4	99.6	4680	1	ECCSGABDG	X90754 E.coli csgG
4		454.4	99.6	10346	1	AE000205	AE000205 Escherich
5		454.4	99.6	15047	1	D90741	D90741 Escherichia
6		395	86.6	1711	1	AF275733	AF275733 Escherich
7		395	86.6	10190	1	AE005315	AE005315 Escherich
8		395	86.6	327773	1	AP002554	AP002554 Escherich
9		382.2	83.8	306358	1	AE016759	AE016759 Escherich
10		340.8	74.7	10370	1	AE015131	AE015131 Shigella
11		340.8	74.7	292504	1	AE016981	AE016981 Shigella
12	c	311.4	68.3	437	1	AF237726	AF237726 Shigella
13		259.2	56.8	5103	1	STAJ2301	AJ002301 Salmonell
14		259.2	56.8	22411	1	AE008749	AE008749 Salmonell
15		257.6	56.5	254050	1	AL627269	AL627269 Salmonell
16	c	257.6	56.5	301983	1	AE016840	AE016840 Salmonell
17		254.4	55.8	456	6	I44909	I44909 Sequence 58
18		254.4	55.8	2067	1	SEU43280	U43280 Salmonella
19		253.6	55.6	1048	1	STAGFBA	AJ000514 Salmonell
20		252.8	55.4	2883	1	ESA515702	AJ515702 Enterobac
21		246.4	54.0	2889	1	CSP515700	AJ515700 Citrobact
22		230.8	50.6	2920	1	CFR515701	AJ515701 Citrobact
23		186	40.8	19201	1	D90742	D90742 Escherichia
24		173.4	38.0	361	6	I44908	I44908 Sequence 56
25		121.8	26.7	230	1	SEU53207	U53207 Salmonella
26		83.8	18.4	1212	1	ECO131756	AJ131756 Escherich
27		78	17.1	78	6	AX814809	AX814809 Sequence
28		72	15.8	72	6	AX814798	AX814798 Sequence
29		66	14.5	66	6	AX814808	AX814808 Sequence
30		43.4	9.5	301214	1	AE016786	AE016786 Pseudomon
31		43.2	9.5	85600	9	AC138647	AC138647 Homo sapi
32	c	42.6	9.3	135259	9	HS127B20	Z83838 Human DNA s
33	c	42.6	9.3	165074	9	AC093519	AC093519 Homo sapi
34		41.6	9.1	181047	2	AC016765	AC016765 Homo sapi
35		41.6	9.1	198599	9	AC108448	AC108448 Homo sapi
36		41.2	9.0	159070	2	AC128114	AC128114 Rattus no
37	c	41.2	9.0	186562	2	AL451079	AL451079 Homo sapi
38	c	41.2	9.0	200368	9	AC096541	AC096541 Homo sapi
39		41.2	9.0	241394	2	AC132789	AC132789 Rattus no
40		40.8	8.9	165074	9	AC093519	AC093519 Homo sapi
41	c	40.6	8.9	236533	2	AC103433	AC103433 Rattus no
42		40.2	8.8	183458	2	AC129554	AC129554 Mus muscu
43	c	40	8.8	180995	9	AC117500	AC117500 Homo sapi
44		40	8.8	187615	2	AC020724	AC020724 Homo sapi
45		39.8	8.7	999	1	AY077581	AY077581 Bradyrhiz

ALIGNMENTS

RESULT 1
AX814811
LOCUS AX814811 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
TITLE Peptides
JOURNAL Patent: WO 03064446-A 15 07-AUG-2003;

Hansa Medical Research Aktiebolag (SE)

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN
```

Query Match 100.0%; Score 456; DB 6; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.6e-122;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

```
RESULT 2
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
```

```
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN
```

Query Match 100.0%; Score 456; DB 1; Length 648;
Best Local Similarity 100.0%; Pred. No. 9e-122;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

```
RESULT 3
ECCSGABDG
```

```
LOCUS       ECCSGABDG              4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match             99.6%;  Score 454.4;  DB 1;  Length 4680;
  Best Local Similarity   99.8%;  Pred. No. 3.5e-121;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 4
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
```

```
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGPYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423
```

```
  Query Match             99.6%;  Score 454.4;  DB 1;  Length 10346;
  Best Local Similarity   99.8%;  Pred. No. 3.9e-121;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

```
RESULT 5
D90741
LOCUS       D90741                15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
```

KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.

FEATURES Location/Qualifiers
source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."

CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS 2906. .3460
/note="ORF_ID:o231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=11
/product="Hypothetical protein HI1543"
/protein_id="BAA35825.1"
/db_xref="GI:4062608"
/translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
RTMAPLTRDAISAMWDELEEDSEE"
CDS 3514. .4053
/note="unnamed protein product; ORF_ID:o231#4"
/codon_start=1
/transl_table=11
/protein_id="BAA35826.1"
/db_xref="GI:4062609"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
LAARSAKTAHREQEIKNKA"
gene complement(4117. .4950)
/gene="csgG"
CDS complement(4117. .4950)
/gene="csgG"
/note="ORF_ID:o231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component CsgG precursor."
/protein_id="BAA35827.1"
/db_xref="GI:4062610"
/translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE

```
                S"
     gene       complement(4977. .5393)
                /gene="csgF"
     CDS        complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene       complement(5418. .5807)
                /gene="csgE"
     CDS        complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
     gene       complement(5812. .6462)
                /gene="csgD"
     CDS        complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene       7217. .7672
                /gene="csgB"
     CDS        7217. .7672

  Query Match             99.6%;  Score 454.4;  DB 1;  Length 15047;
  Best Local Similarity   99.8%;  Pred. No. 4.1e-121;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 6
AF275733
LOCUS       AF275733               1711 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
            CsgA protein (csgA) genes, complete cds.
ACCESSION   AF275733
VERSION     AF275733.1  GI:14039399
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Mutations in the csgD promoter associated with variations in curli
            expression in certain strains of Escherichia coli O157:H7
  JOURNAL   Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
  MEDLINE   21218556
   PUBMED   11319125
REFERENCE   2  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
            Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES             Location/Qualifiers
     source          1. .1711
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="43895 Red Variant"
                     /serotype="O157:H7"
                     /db_xref="taxon:562"
     gene            757. .1212
                     /gene="csgB"
     CDS             757. .1212
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="AAK53211.1"
                     /db_xref="GI:14039400"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1253. .1711
                     /gene="csgA"
     CDS             1253. .1711
                     /gene="csgA"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="AAK53212.1"
                     /db_xref="GI:14039401"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match             86.6%;  Score 395;  DB 1;  Length 1711;
```

```
  Best Local Similarity   92.8%;  Pred. No. 6.3e-104;
  Matches  426;  Conservative    0;  Mismatches   30;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       1253 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 1312

Qy         61 GGTGTTGTTCCTCAGTACGGCGGC---GGCGGTAACCACGGTGGTGGCGGTAATAATAGC 117
              |||||||||||||||||||||||   |||||||||||||||||||||||||||||| |||
Db       1313 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 1372

Qy        118 GGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTG 177
              ||||| ||||| |||||||| ||||| |||||||||| ||||||||||||||||||||||
Db       1373 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 1432

Qy        178 CAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCA 237
              ||| ||||||| |||||||||| | ||||||||||||||| || ||||| ||||||
Db       1433 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 1492

Qy        238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAAC 297
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1493 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 1552

Qy        298 AGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGT 357
              ||||| |||||||||||||||||| ||| | ||| | ||||| |||||||| ||||||
Db       1553 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 1612

Qy        358 GGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAG 417
              || |||||||||| || |||||||||||||||||| ||| ||||||||| ||||||
Db       1613 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 1672

Qy        418 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||
Db       1673 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 1711


RESULT 7
AE005315
LOCUS       AE005315               10190 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
            of 155.
ACCESSION   AE005315 AE005174
VERSION     AE005315.1  GI:12514572
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10190
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            617. .1072
                     /gene="csgB"
                     /note="synonym: Z1675"
     CDS             617. .1072
                     /gene="csgB"
                     /function="structural component; Cell exterior
                     constituents: Surface structures"
                     /note="Residues 1 to 151 of 151 are 100.00 pct identical
                     to residues 1 to 151 of 151 from Escherichia coli K-12
                     Strain MG1655: B1041"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAG55787.1"
                     /db_xref="GI:12514573"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1113. .1571
                     /gene="csgA"
                     /note="synonym: Z1676"
     CDS             1113. .1571
                     /gene="csgA"
                     /function="regulator; Cell exterior constituents: Surface
                     structures"
                     /note="Residues 1 to 152 of 152 are 96.71 pct identical to
                     residues 1 to 151 of 151 from Escherichia coli K-12 Strain
                     MG1655: B1042"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin major subunit, coiled surface structures;
                     cryptic"
                     /protein_id="AAG55788.1"
                     /db_xref="GI:12514574"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
     gene            1630. .1962
                     /gene="csgC"
                     /note="synonym: Z1677"
     CDS             1630. .1962
                     /gene="csgC"
                     /function="putative structure; Not classified"
                     /note="Residues 1 to 110 of 110 are 98.18 pct identical to
                     residues 1 to 110 of 110 from Escherichia coli K-12 Strain
                     MG1655: B1043"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="AAG55789.1"
                     /db_xref="GI:12514575"
                     /translation="MNALLLLAALSSQITFNTTQQGDMYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     gene            2083. .2394
                     /gene="Z1678"
     CDS             2083. .2394
                     /gene="Z1678"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 103 of 103 are 98.05 pct identical to
                     residues 1 to 103 of 103 from Escherichia coli K-12 Strain
                     MG1655: B1044"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55790.1"
                     /db_xref="GI:12514576"
                     /translation="MFRPFLNSLMLGSLFFPFIAIAGSTAQGGVIHFYGQIVEPACDV
                     STQSSPVEMNCPQNGSVPGKTYSSKALMSGNVKNAQIASVKVQYLDKQKKLAVMNIEY
```

```
                N"
gene            2489. .3022
                /gene="Z1679"
CDS             2489. .3022
                /gene="Z1679"
                /function="orf; Unknown function"
                /note="Residues 1 to 177 of 177 are 100.00 pct identical
                to residues 1 to 177 of 177 from Escherichia coli K-12
                Strain MG1655: B1045"
                /codon_start=1
                /transl_table=11
                /product="putative polyprotein"
                /protein_id="AAG55791.1"
                /db_xref="GI:12514577"
                /translation="MKTRIHVVQGDITKLAVDVIVNAANPSLMGGGGVDGAIHRAAGP
                ALLDACLKVRQQQGDCPTGHAVITLAGDLPAKAVVHTVGPVWRGGEQNEDQLLQDAYL
                NSLRLVAANSYTSVAFPAISTGVYGYPRAAAAEIAVKTVSEFITRHALPEQVYFVCYD
                EENAHLYERLLTQQGDE"
gene            2964. .4445
                /gene="ymdC"
                /note="synonym: Z1680"
CDS             2964. .4445
                /gene="ymdC"
                /function="putative enzyme; Not classified"
                /note="Residues 1 to 493 of 493 are 99.59 pct identical to
                residues 1 to 493 of 493 from Escherichia coli K-12 Strain
                MG1655: B1046"
                /codon_start=1
                /transl_table=11
                /product="putative synthase"
                /protein_id="AAG55792.1"
                /db_xref="GI:12514578"
                /translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPGQCGLFPL
                EKSLDAFAARYRLAEMSEHTLDVQYYIWQDDMSGRLLFSALLAAAKRGVRVRLLLDDN
                NTPGLDDILRLLDSHPRIEVRLFNPFSFRLLRPLGYITDFSRLNRRMHNKSFTVDGVV
                TLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWYCKSVSPLQQVLDV
                PEGEMADRIELPASWHNDAMTHRYLRKMESSPFINHLVDGTLPLIWAKTRLLSDDPAK
                GEGKAKRHSLLPQRLFDIMGSPSERIDIISSYFVPTRAGVAQLLRMVRKXVKIAILTN
                SLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREQSSTLHDRGITGNSGASLHAKT
                FSIDGKTVFIGSFNFDPRSTLLNTEMGFVIESETLAQLIDKRFIQSQYDAAWQLRLDR
                WGRINWVDRHAKKEIVLKKEPATSFWKRVMVRLASILPVEWLL"
gene            complement(4453. .5610)
                /gene="ymdD"
                /note="synonym: Z1681"
CDS             complement(4453. .5610)
                /gene="ymdD"
                /function="orf; Unknown function"
                /note="Residues 1 to 385 of 385 are 99.22 pct identical to
                residues 1 to 385 of 385 from Escherichia coli K-12 Strain
                MG1655: B1047"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG55793.1"
                /db_xref="GI:12514579"
                /translation="MNPVPAQREYFXDSIRAWLMLLGIPFHISLIYSSHTWHVNSAEP
                SLWLTLFNDFIHSFRMQVFFVISGYFSYMLFLRYPLKKWWKVRVERVGIPMLTAIPLL
                TLPQFIMLQYVKGKAESWPGLSLYDKYNTLAWELISHLWFLLVLVVMTTLCVWIFKRI
                RNNLENSDKTNKKFSMVKLSVIFLCLGIGYAVIRRTIFIVYPPILSNGTFNFIVMQTL
                FYLPFFILGALAFIFPHLKALFTTPSRGCTLAAALAFVAYLLNQRYGSGDAWMYETES
                VITMVLGLWMVNVVFSFGHRLLNFQSARVTYFVNASLFIYLVHHPLTLFFGAYITPHI
                TSNWLGFLCGLIFVVGIAIILYEIHLRIPLLKFLFSGKPVVKRENDKAPAR"
gene            6003. .7538
                /gene="mdoG"
                /note="synonym: Z1683"
CDS             6003. .7538
                /gene="mdoG"
                /function="enzyme; Osmotic adaptation"
                /note="Residues 1 to 511 of 511 are 99.80 pct identical to
                residues 1 to 511 of 511 from Escherichia coli K-12 Strain
                MG1655: B1048"
                /codon_start=1
                /transl_table=11
                /product="periplasmic glucans biosynthesis protein"
                /protein_id="AAG55794.1"
                /db_xref="GI:12514580"
                /translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAQSLAGKGYEAPK
                SNLPSVFRDMKYADYQQIQFNHDKAYWNNLKTPFKLEFYHQGMYFDTPVKINEVTATA
                VKRIKYSPDYFTFGDVQHDKDTVKDLGFAGFKVLYPINSKDKNDEIVSMLGASYFRVI
                GAGQVYGLSARGLAIDTALPSGEEFPRFKEFWIERPKPTDKRLTIYALLDSPRATGAY
                KFVVMPGRDTVVDVQSKIYLRDKVGKLGVAPLTSMFLFGPNQPSPANNYRPELHDSNG
                LSIHAGNGEWIWRPLNNPKHLAVSSFSMENPQGFGLLQRGRDFSRFEDLDDRYDLRPS
                AWVTPKGEWGKGSVELVEIPTNDETNDNIVAYWTPDQLPEPGKEMNFKYTITFSRDED
                KLHAPDNAWVQQTRRSTGDVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDTPVTAQTS
                IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDAKKTTEMRAALVNADQTLSETWSYQLP
                ANE"
gene            7531. .10074
                /gene="mdoH"
                /note="synonym: Z1684"
CDS             7531. .10074
                /gene="mdoH"
                /function="enzyme; Osmotic adaptation"
```

Query Match 86.6%; Score 395; DB 1; Length 10190;
Best Local Similarity 92.8%; Pred. No. 8.1e-104;
Matches 426; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

```
Qy       1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1113 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 1172

Qy      61 GGTGTTGTTCCTCAGTACGGCGGC---GGCGGTAACCACGGTGGTGGCGGTAATAATAGC 117
          |||||||||||||||||||||||   |||||||||||||||||||||||||||||| |||
Db    1173 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 1232

Qy     118 GGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTG 177
          ||||| ||||| |||||||| ||||| |||||||||||| ||||||||||||||||||||||
Db    1233 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 1292

Qy     178 CAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCA 237
          ||| ||||||| |||||||||| | |||||||||||||||| || ||||| ||||||
Db    1293 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 1352

Qy     238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAAC 297
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    1353 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 1412

Qy     298 AGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGT 357
          ||||| |||||||||||||||||| ||| | ||| | ||||| |||||||| ||||||
Db    1413 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 1472

Qy     358 GGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAG 417
          || |||||||||| || |||||||||||||||||| ||| |||||||||| ||||||
Db    1473 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 1532

Qy     418 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
          |||||||||||||||||||||||||||||||||||||||
Db    1533 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 1571
```

```
RESULT 8
AP002554
LOCUS       AP002554              327773 bp    DNA     linear   BCT 21-DEC-2002
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION   AP002554 BA000007
VERSION     AP002554.1  GI:13360491
KEYWORDS    .
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
```

```
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
   PUBMED   10734605
REFERENCE   2
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
   PUBMED   11108008
REFERENCE   3
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
   PUBMED   11111050
REFERENCE   4
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
   PUBMED   11258796
REFERENCE   5  (bases 1 to 327773)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .327773
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            150. .2258
                     /gene="ECs1032"
     CDS             150. .2258
                     /gene="ECs1032"
                     /note="similar to YCBY_ECOLI gi|1787181 percent identity
                     99 in 702 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative oxidoreductase"
                     /protein_id="BAB34455.1"
                     /db_xref="GI:13360492"
                     /translation="MNSLFASTARGLEELLKTELENLGAVECQVVQGGVHFKGDTRLV
                     YQSLMWSRLASRIMLPLGECKVYSDLDLYLGVQAINWTEMFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRDAPDIRVNVWLHKETASIALDLSGDGLH
                     LRGYRDRAGIAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMLATDRAPGL
                     HRGRWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIQRARTNAR
                     LAGIGELITFEVKDVAQLTNPLPKEPYGTVLSNPPYGERLDSEPALIALHSLLGRIMK
                     NQFGGWNLSLFSASPDLLSCLQLRADKQYKAKNGPLDCVQKNYHVAESTPDSKPAMVA
                     EDYANRLRKNLKKFEKWARQEGIECYRLYDADLPEYNVAVDRYADWVVVQEYAPPKTI
                     DAHKARQRLFDIIAATISVLGIAPNKLVLKTRERQKGKNQYQKLGEKGEFLEVTEYNA
                     HLWVNLTDYLDTGLFLDHRIARRMLGQMSKGKDFLNLFSYTGSATVHAGLGGARSTTT
                     VDMSRTYLEWAERNLRLNGLTGRAHRLIQADCLAWLREANEQFDLIFIDPPTFSNSKR
                     MEDAFDVQRDHLALMKDLKRLLRAGGTIMFSNNKRGFRMDLDGLAKLGLKAQEITQKT
                     LSQDFARNRQIHNCWLITAA"
     gene            2270. .4177
                     /gene="ECs1033"
     CDS             2270. .4177
                     /gene="ECs1033"
                     /note="similar to UUP_ECOLI gi|1787182 percent identity 99
                     in 635 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative ATP-binding component of a transport
                     system"
                     /protein_id="BAB34456.1"
                     /db_xref="GI:13360493"
                     /translation="MSLISMHGAWLSFSDAPLLDNAELHIEDNERVCLVGRNGAGKST
                     LMKILNREQGLDDGRIIYEQDLIVARLQQDPPRNVEGSVYDFVAEGIEEQAEYLKRYH
                     DISRLVMNDPSEKNLNELAKVQEQLDHHNLWQLENRINEVLAQLGLDPNVALSSLSGG
                     WLRKAALGRALVSNPRVLLLDEPTNHLDIETIDWLEGFLKTFNGTIIFISHDRSFIRN
                     MATRIVDLDRGKLVTYPGNYDQYLLEKEEALRVEELQNAEFDRKLAQEEVWIRQGIKA
                     RRTRNEGRVRALKAMRRERGERREVMGTAKMQVEEASRSGKIVFEMEDVCYQVDGKQL
                     VKDFSAQVLRGDKIALIGPNGCGKTTLLKLMLGQLQADSGRIHVGTKLEVAYFDQHRA
                     ELDPDKTVMDNLAEGKQEVMVNGKPRHVLGYLQDFLFHPKRAMTPVRALSGGERNRLL
                     LARLFLKPSNLLILDEPTNDLDVETLELLEELIDSYQGTVLLVSHDRQFVDNTVTECW
                     IFEGGGKIGRYVGGYHDARGQQEQYVALKQPAVKKNEEPAAPKAETVKRSSSKLSYKL
                     QRELEQLPQLLEDLEAKLEALQTQVADASFFSQPHEQTQKVLADMAAAEQELEQAFER
                     WEYLEALKNGG"
     gene            4307. .5560
                     /gene="ECs1034"
     CDS             4307. .5560
                     /gene="ECs1034"
                     /note="similar to PQIA_ECOLI gi|1787183 percent identity
                     99 in 417 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="paraquat-inducible protein A"
                     /protein_id="BAB34457.1"
                     /db_xref="GI:13360494"
                     /translation="MCEHHHAAKHILCSQCDMLVALPRLEHGQKAACPRCGTTLTVAW
                     DAPRQRPTAYALAALFMLLLSNLFPFVNMNVAGVTSEITLLEIPGVLFSEDYASLGTF
                     FLLFVQLVPAFCLITILLLVNRAELPVRLKEQLARVLFQLKTWGMAEIFLAGVLVSFV
                     KLMAYGSIGVGSSFLPWCLFCVLQLRAFQCVDRRWLWDDIAPMPELRQPLKPGVTGIR
                     QGLRSCSCCTAILPADEPVCPRCGTKGYVRRRNSLQWTLALLVTSIMLYLPANILPIM
                     VTDLLGSKMPSTILAGVILLWSEGSYPVAAVIFLASIMVPTLKMIAIAWLCWDAKGHG
                     KRDSERMHLIYEVVEFVGRWSMIDVFVIAVLSALVRMGGLMSIYPAMGALMFALVVIM
                     TMFSAMTFDPRLSWDRQPESEHEES"
     gene            5565. .7205
                     /gene="ECs1035"
     CDS             5565. .7205
                     /gene="ECs1035"
                     /note="similar to PQIB_ECOLI gi|1787184 percent identity
                     99 in 546 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="paraquat-inducible protein B"
                     /protein_id="BAB34458.1"
                     /db_xref="GI:13360495"
                     /translation="MESNNGEAKIQKVKNWSPVWIFPIVTALIGAWVLFYHYSHQGPE
                     VTLITANAEGIEGGKTTIKSRSVDVGVVESATLADDLTHVEIKARLNSGMEKLLHKDT
                     VFWVVKPQIGREGISGLGTLLSGVYIELQPGAKGSKMDKYDLLDSPPLAPPDAKGIRV
                     VLDSKKAGQLSPGDPVLFRGYRVGSVETSTFDTQKRNISYQLFINAPYDRLVTSNVRF
                     WKDSGIAVDLTSAGMRVEMGSLTTLLSGGVSFDVPEGLDLGQPVAPKTAFVLYDDQKS
                     IQDSLYTDHIDYLMFFKDSVRGLQPGAPVEFRGIRLGTVSKVPFFAPNMRQTFNDDYR
                     IPVLIRIEPERLKMQLGENADVVEHLGELLKRGLRGSLKTGNLVTGALYVDLDFYPNT
                     PAITGIREFNGYQIIPTVSGGLAQIQQRLMEALDKINKLPLNPMIEQATSTLSESQRT
                     MKNLQTTLDSMNKILASQSMQQLPTDMQSTLRELNRSMQGFQPGSAAYNKMVADMQRL
```

```
                DQVLRELQPVLKTLNEKSNALVFEAKDKKDPEPKRAKQ"
     gene       7217. .7765
                /gene="ECs1036"
     CDS        7217. .7765
                /gene="ECs1036"
                /note="similar to YMBA_ECOLI gi|1787185 percent identity
                98 in 182 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB34459.1"
                /db_xref="GI:13360496"
                /translation="MTIAALWLAGCSSGEINKNYYQLPVVQSGTQSTASQGNRLLWVE
                QVAVPDYLAGNGVVYQTSDVKYVIANNNLWASPLDQQLRNTLVANLSTQLPGWVVASQ
                PLGSAQDTLNVTVTEFNGRYDGKVIVSGEWLLNHQGQLIKRPFRLEGVQTQDGYDEMV
                KVLAGVWSQEAASIAQEIKRLP"
     gene       8021. .8188
                /gene="ECs1037"
     CDS        8021. .8188
                /gene="ECs1037"
                /note="identical to RMF_ECOLI gi|1787186 (Conserved in
                E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="ribosome modulation factor"
                /protein_id="BAB34460.1"
                /db_xref="GI:13360497"
                /translation="MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGW
                REAMADRVVMA"
     gene       complement(8258. .8776)
                /gene="ECs1038"
     CDS        complement(8258. .8776)
                /gene="ECs1038"
                /note="similar to FABA_ECOLI gi|1787187 percent identity
                100 in 172 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11

  Query Match             86.6%;  Score 395;  DB 1;  Length 327773;
  Best Local Similarity   92.8%;  Pred. No. 1.3e-103;
  Matches  426;  Conservative    0;  Mismatches   30;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     326421 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 326480

Qy         61 GGTGTTGTTCCTCAGTACGGCGGC---GGCGGTAACCACGGTGGTGGCGGTAATAATAGC 117
              ||||||||||||||||||||||||   |||||||||||||||||||||||||||||| |||
Db     326481 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 326540

Qy        118 GGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTG 177
              ||||| ||||| |||||||| ||||| |||||||||| ||||||||||||||||||||||
Db     326541 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 326600

Qy        178 CAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCA 237
              ||| ||||||| |||||||||| | |||||||||||||||| || ||||| ||||||
Db     326601 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 326660

Qy        238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAAC 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     326661 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 326720

Qy        298 AGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGT 357
              ||||| |||||||||||||||||| ||| | ||| | ||||| |||||||| ||||||
Db     326721 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 326780

Qy        358 GGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAG 417
              || |||||||||| || |||||||||||||||||||| ||| |||||||||| ||||||
Db     326781 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 326840

Qy        418 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||
Db     326841 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 326879


RESULT 9
AE016759
LOCUS       AE016759               306358 bp    DNA     linear   BCT 24-DEC-2002
DEFINITION  Escherichia coli CFT073 section 5 of 18 of the complete genome.
ACCESSION   AE016759 AE014075
VERSION     AE016759.1  GI:26107527
KEYWORDS    .
SOURCE      Escherichia coli CFT073
  ORGANISM  Escherichia coli CFT073
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Extensive Mosiac Structure Revealed by the Complete Genome Sequence
            of Uropathogenic Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
   PUBMED   12471157
REFERENCE   2  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .306358
                     /organism="Escherichia coli CFT073"
                     /mol_type="genomic DNA"
                     /strain="CFT073"
                     /db_xref="taxon:199310"
     gene            complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
     CDS             complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
                     /note="Residues 27 to 1231 of 1245 are 80.16 pct identical
                     to residues  1 to 1209 of 1218 from GenPept.129 :
                     >emb|CAD05883.1| (AL627276) putative ABC transporter
                     protein [Salmonella enterica subsp. enterica serovar
                     Typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP binding cassette (ABC) transporter homolog"
                     /protein_id="AAN79727.1"
                     /db_xref="GI:26107528"
                     /translation="MRDRSVLAFLRNDPASGSYFTDGGGLMPANHTPTPAQSWIVRLA
                     RVCWERKKLSVIVVVASVSTILLAALTPLLTRQAVNDALAGNPARLPWLACGLLLIAF
                     FDFIGNYVRRGYAGMLSLWVQHTLRGRVFDSIQKLDGAGQDALRTGQVISRTNSDLQQ
                     VHTLLQMCPVPLAVFTYYIAGIAVMLWMSPAMTLIVVCVLVCLAITALRARRRVFAQT
                     GMASDQLANLTEHIREVLAQISVVKSCVAEMRETHWLDRQSRQIVRVRIGAVISQAMP
                     GATMLALPVLGQIVLLCYGGWSVMHGRIDLGTFVAFASFLAMLTGPTRVLASFLVIAQ
                     RTQASVERVFALIDTRSQMEDGTESINSQVVGLELENMSFDYHHGDRHILSDISFSLR
                     AGETVAVVGASGSGKSTLLMLLARFYDPCSGKIWLNTSEGRQNLRDIRLEALRRRVGI
                     VFEDAFLFAGTVAENIAYGHPQATADDIRRAAAAAGASDFINALPKGFDSLLTERGTN
                     LSGGQRQRIALARALITAPDVLILDDTTSAVDAVTEAEINTALGRYADEGHMLLVIAR
                     RRSTLQLASRVVVLDKGRMVDTGTPAELEARCPAFRALMTGDSDFLATSHNSHNELWP
                     AEPATQDDVTDTGDKGFVARMTRVPENAVQQALAGKGRKVTSLLKPVAWMFVIAALLI
                     ALDSAAGVGVLILLQHGIDSGVAAGDMSTIGLCALLALCLVIVGWCSYSLQTVFAARA
                     AESVQHSVRLRSFGHMLRLGLPWHEKHADSRLTRMTVDVDSLARFLQNGLAGAATSLV
                     TMFAIAATMFWLDPILALTALSAVPVAALATMIYRRLSTPAYAQARLEIGKVNSTLQE
                     KVSGLRVVQSHGQQELEGARLRALSERFRATRVRAQKYLAVYFPFLTFCTEASYAAVL
                     LVGASQVAAGEMTAGVLAAFFLLLGQFYGPVQQLSGIVDAWQQATASGKHIDELLATE
                     GTENLGSSSVLPVTGALHLDEVTFSYPDSHEPALNKLTLTIPEGMVVAVVGRSGAGKS
```

```
                TLIKLIAGLYFPTHGNIRIGVQMLDDASLTEYRRQIGLVDQDVALFSSDIAENIRYSR
                PSATNEDVEIASQRAGLYEMVCNLPQGFRTPVNNGGADLSAGQRQLIALARAQLANAH
                ILLLDEATSCLDRTSEERLMSSLTDVVHAGKHSALIVAHRLTTAQRCDLIAVIDKGLL
                AEYGTHEQLLSAGGLYTRLWHDSVSSTALHRQHNMKEETPG"
gene            complement(3882. .5045)
                /gene="iroB"
                /locus_tag="c1254"
CDS             complement(3882. .5045)
                /gene="iroB"
                /locus_tag="c1254"
                /note="Escherichia coli O157:H7 ortholog: z1190"
                /codon_start=1
                /transl_table=11
                /product="Putative glucosyltransferase"
                /protein_id="AAN79728.1"
                /db_xref="GI:26107529"
                /translation="MRRLPDLDRQAERDFLMRILFVGPPLYGLLYPVLSLAQAFRVNG
                HEVLIASGGQFAQKAAEAGLVVFDAAPGLDSEAGYRHHEAQRKKSNIGTQMGNFSFFS
                EEMADHLVEFAGHWRPDLIIYPPLGVIGPLIAAKYDIPVVMQTVGFGHTPWHIRGVTR
                SLTDAYRRHNVGATPRDMAWIDVTPPSMSILENDGEPIIPMQYVPYNGGAVWEPWWER
                RPDRKRLLVSLGTVKPMVDGLDLIAWVMDSASEVDAEIILHISANARSDLRSLPSNVR
                LVDWIPMGVFLNGADGFIHHGGAGNTLTALHAGIPQIVFGQGADRPVNARVVAERGCG
                IIPGDVGLSSNMINAFLNNRSLRKASEEVAAEMAAQPCPGEVAKSLITMVQKG"
gene            5535. .5834
                /locus_tag="c1255"
CDS             5535. .5834
                /locus_tag="c1255"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79729.1"
                /db_xref="GI:26107530"
                /translation="MKPTLLQTVENICLQILFVLFLFTGTLIDVEALVSCSSVAKNVQ
                RTPASSVSSSVYSASVSRTAADIMLLSNNPLFAPGQVSFFGCSLAASGIHVTILE"
repeat_region   5867. .6577
                /insertion_seq="IS1541A-like"
gene            5961. .6470
                /locus_tag="c1256"
CDS             5961. .6470
                /locus_tag="c1256"
                /note="Residues 18 to 169 of 169 are 94.07 pct identical
                to residues  1 to 152 of 152 from SwissProt.40 :
                >sp|Q57334|T200_SALTY Transposase for insertion sequence
                element IS200"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79730.1"
                /db_xref="GI:26107531"
                /translation="MRSGNCKCQTRNQKGVPMGNEKSLAHTRWNCKYHIVFAPKYRRQ
                VFYREKRRAIGCILRKLCEWKSVRILEAECCADHIHMLVEIPPKMSVSGFMGYLKGKS
                SLMPYEQFGDLKFKYRNREFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPY
                PGSPFTGRK"
repeat_region   complement(6578. .6940)
                /note="incomplete"
                /insertion_seq="IS4"
gene            complement(6598. .6960)
                /locus_tag="c1257"
CDS             complement(6598. .6960)
                /locus_tag="c1257"
                /function="IS, phage, Tn"
                /note="Residues 8 to 120 of 120 are 98.23 pct identical to
                residues 316 to 428 of 442 from MG1655 : b4278"
                /codon_start=1
                /transl_table=11
                /product="Putative conserved protein"
                /protein_id="AAN79731.1"
                /db_xref="GI:26107532"
                /translation="MPRQPDQGYREIKQTMQLSRLTLRSKKPELVEQELWGVLLAYNL
                VRYQMIKMAEHLKGYWPNQLSFSESCGMVMRMLMTLQGASPGRIPELMRDLASMGQLV
                KLPTRRETAFPRVVKERP"
gene            7651. .8013
                /locus_tag="c1258"
CDS             7651. .8013
                /locus_tag="c1258"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79732.1"
                /db_xref="GI:26107533"
                /translation="MLGACFYSWGTSFYYGFPWWGAELVLMMWPEACFMLLQSSLYSL
                LDGEQVLFSFWELNKKQYEGFEFYQIFSGNIVASCSSCSGVSSNSSAHYARCTDFMRH
                YSLNNYAACQLGRQACLS"
gene            7835. .8269
                /locus_tag="c1259"
CDS             7835. .8269
                /locus_tag="c1259"
                /note="Residues 1 to 139 of 144 are 50.71 pct identical to
                residues  104 to 242 of 258 from GenPept.129 :
                >gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
                [Shigella flexneri]"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79733.1"
                /db_xref="GI:26107534"
                /translation="MKDLSFIRFFLAILLLLVPPALEFPVIHRHITPGALTLCVITAL
                IITLLASLGGRLVSVKCISEISFIRRHRIECMMAGFMIYFWTFSLIAGWYKPQFKKEY
                QTLHYENARYYVLARYDGRLVLSQSYRSGSRKFVIINGGHRL"
gene            8477. .8553
                /locus_tag="c5534"
tRNA            8477. .8553
                /locus_tag="c5534"
                /product="tRNA-Arg"
                /function="RNA; tRNA"
gene            complement(8832. .9290)
                /locus_tag="c1260"
CDS             complement(8832. .9290)
                /locus_tag="c1260"
                /function="putative enzyme; Transposases Insertion
                Sequence Associated"
                /note="Escherichia coli O157:H7 ortholog: z4324"
                /codon_start=1
                /transl_table=11
                /product="Putative Transposase"
                /protein_id="AAN79734.1"
                /db_xref="GI:26107535"
                /translation="MASAADARRIVSHYERRWLIEEYHKAWKSGGTCVESLRMQTRDN
                LERMVVIKAFIAVRMLGLRQEGISEETQNDSCKKILTPTEWKLLWVKLEGKQLPSQTP
                TLKWACLKLGRWHDSKRTGRPGWVVMWDGWFRLQDMVEGYPVMKSLDQEI"
gene            complement(9302. .9667)
                /locus_tag="c1261"
CDS             complement(9302. .9667)
                /locus_tag="c1261"
                /note="Residues 1 to 35 of 121 are 97.14 pct identical to
                residues  59 to 93 of 316 from GenPept.129 :
                >gb|AAK18437.1|AF348706_126 (AF348706) putative
                transposase [Shigella flexneri]"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79735.1"
                /db_xref="GI:26107536"
                /translation="MAERLGEIQKRVITVCDREADIWHYLHYKVSHGQRFVVRTAQNS
```

```
Query Match             83.8%;  Score 382.2;  DB 1;  Length 306358;
Best Local Similarity   91.1%;  Pred. No. 7.1e-100;
Matches  418;  Conservative    0;  Mismatches   38;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      45269 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 45328

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGT---AACCACGGTGGTGGCGGTAATAATAGC 117
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||| |||
Db      45329 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTGGCAACCACGGTGGTGGCGGTAATAACAGC 45388
```

```
Qy        118 GGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTG 177
              || || ||||| |||||||| ||||||||||||||||||||||||||||| ||||||| |
Db      45389 GGTCCGAATTCAGAGCTGAATATTTACCAGTACGGTGGCGGTAACTCTGCTCTTGCTCAG 45448

Qy        178 CAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCA 237
              ||| |||| |||||||||||||||  |||| |||||||||| |||||||||||||| |||
Db      45449 CAAGCTGACGCCCGTAACTCTGATCTGACCATTACCCAGCACGGCGGCGGTAATGGCGCA 45508

Qy        238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAAC 297
              ||||| || || || || ||||||||||||||||||||||| || ||||| ||||| |||
Db      45509 GATGTGGGCCAAGGTTCTGATGACAGCTCAATCGATCTGACTCAGCGTGGTTTCGGCAAC 45568

Qy        298 AGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGT 357
              |||||||||||||||||||||| || ||| |||||   ||||| ||||||||||||||||
Db      45569 AGCGCTACTCTTGATCAGTGGAATGGTAAAGATTCTACTATGACTGTTAAACAGTTCGGT 45628

Qy        358 GGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAG 417
              || || |||||||||| |||||||||||||||||||||||  ||| ||||| ||||||
Db      45629 GGCGGTAACGGTGCTGCTGTTGACCAGACTGCATCTAACTCCAGCGTTAACGTCACTCAG 45688

Qy        418 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||
Db      45689 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 45727
```

RESULT 10
AE015131
LOCUS AE015131 10370 bp DNA linear BCT 18-OCT-2002
DEFINITION Shigella flexneri 2a str. 301 section 94 of 412 of the complete genome.
ACCESSION AE015131 AE005674
VERSION AE015131.1 GI:24051313
KEYWORDS .
SOURCE Shigella flexneri 2a str. 301
ORGANISM Shigella flexneri 2a str. 301
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 10370)
AUTHORS Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H., Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L., Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
TITLE Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157
JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)
PUBMED 12384590
REFERENCE 2 (bases 1 to 10370)
AUTHORS Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y., Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R. China

```
FEATURES             Location/Qualifiers
     source          1. .10370
                     /organism="Shigella flexneri 2a str. 301"
                     /mol_type="genomic DNA"
                     /strain="301"
                     /serotype="2a"
                     /db_xref="taxon:198214"
     tRNA            complement(127. .214)
                     /product="tRNA-Ser"
                     /note="anticodon: GGA"
     gene            410. .1387
                     /gene="ycdW"
                     /locus_tag="SF1027"
                     /pseudo
     CDS             410. .1387
                     /gene="ycdW"
                     /locus_tag="SF1027"
                     /note="similar to Escherichia coli K12 putative
                     dehydrogenase gi: 1787270 (326 aa). BLAST with identity of
                     99% in 326 aa. This CDS contains an in-frame stop codon.
                     The sequence has been checked and is believed to be
                     correct."
                     /pseudo
                     /codon_start=1
                     /transl_table=11
     gene            1442. .2179
                     /gene="ycdX"
                     /locus_tag="SF1028"
     CDS             1442. .2179
                     /gene="ycdX"
                     /locus_tag="SF1028"
                     /note="Residues 1 to 245 of 245 are 99 pct identical to
                     residues 1 to 245 of a 245 aa protein from Escherichia
                     coli O157:H7 EDL933 ref: NP_287168.1"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, conserved hypothetical protein"
                     /protein_id="AAN42652.1"
                     /db_xref="GI:24051314"
                     /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                     EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                     PVFAPHDKATNTQAMISTIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                     SFLHSRKGSEDNCRAVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPLER
                     ILNVSPRRLLNFLESRGMAPIAEFADL"
     gene            2203. .2757
                     /gene="ycdY"
                     /locus_tag="SF1029"
     CDS             2203. .2757
                     /gene="ycdY"
                     /locus_tag="SF1029"
                     /note="Residues 1 to 184 of 184 are 99 pct identical to
                     residues 1 to 184 of a 184 aa protein from Escherichia
                     coli O157:H7 EDL933 ref: NP_287169.1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative oxidoreductase component"
                     /protein_id="AAN42653.1"
                     /db_xref="GI:24051315"
                     /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                     DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEDATEAEVRAFLSERGMLL
                     ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                     RTMAPLTRDAISAMWDELEEDSEE"
     gene            2811. .3350
                     /gene="ycdZ"
                     /locus_tag="SF1030"
     CDS             2811. .3350
                     /gene="ycdZ"
                     /locus_tag="SF1030"
                     /note="Residues 1 to 179 of 179 are 98 pct identical to
                     residues 1 to 179 of a 179 aa protein from Escherichia
                     coli K12 ref: NP_415554.1"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, conserved hypothetical protein"
                     /protein_id="AAN42654.1"
                     /db_xref="GI:24051316"
                     /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                     SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                     IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLVFGYAMKNSGLW
                     LAARSAKTAHREQEIKNKA"
     gene            complement(3414. .4189)
                     /gene="csgG"
                     /locus_tag="SF1031"
                     /pseudo
     CDS             complement(3414. .4189)
                     /gene="csgG"
                     /locus_tag="SF1031"
                     /note="similar to Escherichia coli K12 curli production
```

```
                assembly-transport component, 2nd curli operon gi: 1787274
                (278 aa). BLAST with identity of 97% in 257 aa. This CDS
                contains an in-frame stop codon. The sequence has been
                checked and is believed to be correct."
                /pseudo
                /codon_start=1
                /transl_table=11
     gene       complement(4454. .5662)
                /locus_tag="SF1032"
     CDS        complement(4454. .5662)
                /locus_tag="SF1032"
                /note="Residues 1 to 402 of 402 are 100 pct identical to
                residues 1 to 402 of a 402 aa protein from Plasmid R100
                ref: NP_052925.1"
                /codon_start=1
                /transl_table=11
                /product="transposase of Tn10"
                /protein_id="AAN42655.1"
                /db_xref="GI:24051317"
                /translation="MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTLTLTEL
                GRNLPTKARTKHNIKRIDRLLGNRHLHKERLAVYRWHASFICSGNTMPIVLVDWSDIR
                EQKRLMVLRASVALHGRSVTLYEKAFPLSEQCSKKAHDQFLADLASILPSNTTPLIVS
                DAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAENWKPISNLHDMSSSHSKTLGYKR
                LTKSNPISCQILLYKSRSKGRKNQRSTRTHCHHPSPKIYSASAKEPWILATNLPVEIR
                TPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWL
                AGVHAQKQGWDKHFQANTVRNRNVLSTVRLGMEVLRHSGYTITREDSLVAATLLTQNL
                FTHGYVLGKL"
     gene       complement(5995. .6384)
                /gene="csgE"
                /locus_tag="SF1033"
     CDS        complement(5995. .6384)
                /gene="csgE"
                /locus_tag="SF1033"
                /note="Residues 1 to 129 of 129 are 98 pct identical to
                residues 1 to 129 of a 129 aa protein from Escherichia
                coli O157:H7 EDL933 ref: NP_287173.1"
                /codon_start=1
                /transl_table=11
                /product="curli production assembly-transport component,
                2nd curli operon"
                /protein_id="AAN42656.1"
                /db_xref="GI:24051318"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDSEKTVVFAL
                IQTEEALNRRQINQALLSTDDLAHDEF"
     gene       complement(6389. .7069)
                /gene="csgD"
                /locus_tag="SF1034"
     CDS        complement(6389. .7069)
                /gene="csgD"
                /locus_tag="SF1034"
                /note="Residues 11 to 226 of 226 are 100 pct identical to
                residues 1 to 216 of a 216 aa protein from Escherichia
                coli O157:H7 EDL933 ref: NP_287174.1"
                /codon_start=1
                /transl_table=11
                /product="putative 2-component transcriptional regulator
                for 2nd curli operon"
                /protein_id="AAN42657.1"
                /db_xref="GI:24051319"
                /translation="MRSIKKWGFIMFNEVHSIHGHTLLLITKPSLQATALLQHLKQSL
                AITGKLHNIQRSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPED
                YPYRDIENWPHINGVFYAMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNS
                TESALLTHREKEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVS
                WANDNLRR"
     gene       7766. .8248
                /gene="csgB"
                /locus_tag="SF1035"
     CDS        7766. .8248
                /gene="csgB"
                /locus_tag="SF1035"
                /note="Residues 10 to 160 of 160 are 99 pct identical to
                residues 1 to 151 of a 151 aa protein from Escherichia
                coli O157:H7 EDL933 ref: NP_287175.1"
                /codon_start=1
                /transl_table=11
                /product="minor curlin subunit precursor, similar ro CsgA"
                /protein_id="AAN42658.1"
                /db_xref="GI:24051320"
                /translation="MYDQVQGDNMKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAV
                NELSKSSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYI
                DQAGSANDASISQGAYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR

  Query Match             74.7%;  Score 340.8;  DB 1;  Length 10370;
  Best Local Similarity   98.0%;  Pred. No. 5.2e-88;
  Matches  345;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
              | |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       9538 CAGTAATAATAGCGGCCCAAATTCAGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 9597

Qy        165 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9598 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 9657

Qy        225 CGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9658 TGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 9717

Qy        285 TGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGT 344
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| ||
Db       9718 TGGCTTCGGTAACAGCGCCACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACTGT 9777

Qy        345 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGT 404
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       9778 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCGAACTCCTCCGT 9837

Qy        405 CAACGTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       9838 CAACGTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTATTAA 9889


RESULT 11
AE016981
LOCUS       AE016981               292504 bp    DNA     linear   BCT 22-APR-2003
DEFINITION  Shigella flexneri 2a str. 2457T section 4 of 16 of the complete
            genome.
ACCESSION   AE016981 AE014073
VERSION     AE016981.1  GI:30040616
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 2457T
  ORGANISM  Shigella flexneri 2a str. 2457T
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Complete Genome Sequence and Comparative Genomics of Shigella
            flexneri Serotype 2a Strain 2457T
  JOURNAL   Infect. Immun. 71 (5), 2775-2786 (2003)
   PUBMED   12704152
REFERENCE   2  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .292504
                     /organism="Shigella flexneri 2a str. 2457T"
                     /mol_type="genomic DNA"
                     /strain="2457T"
```

```
                /serotype="2a"
                /db_xref="taxon:198215"
     gene       complement(223. .441)
                /gene="infA"
                /locus_tag="S0884"
     CDS        complement(223. .441)
                /gene="infA"
                /locus_tag="S0884"
                /function="factor; Proteins - translation and
                modification"
                /note="residues 1 to 72 of 72 are 100.00 pct identical to
                residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
                /codon_start=1
                /transl_table=11
                /product="protein chain initiation factor IF-1"
                /protein_id="AAP16348.1"
                /db_xref="GI:30040617"
                /translation="MAKEDNIEMQGTVLETLPNTMFRVELENGHVVTAHISGKMRKNY
                IRILTGDKVTVELTPYDLSKGRIVFRSR"
     gene       complement(726. .1430)
                /gene="aat"
                /locus_tag="S0885"
     CDS        complement(726. .1430)
                /gene="aat"
                /locus_tag="S0885"
                /function="enzyme; Aminoacyl tRNA synthetases, tRNA
                modification"
                /note="residues 1 to 234 of 234 are 100.00 pct identical
                to residues 1 to 234 of 234 from Escherichia coli K-12 :
                B0885"
                /codon_start=1
                /transl_table=11
                /product="leucyl, phenylalanyl-tRNA-protein transferase"
                /protein_id="AAP16349.1"
                /db_xref="GI:30040618"
                /translation="MRLVQLSRHSIAFPSPEGALREPNGLLALGGDLSPARLLMAYQR
                GIFPWFSPGDPILWWSPDPRAVLWPESLHISRSMKRFHKRSPYRVTMNYAFGQVIEGC
                ASDREEGTWITRGVVEAYHRLHELGHAHSIEVWREDELVGGMYGVAQGTLFCGESMFS
                RMENASKTALLVFCEEFIGHGGKLIDCQVLNDHTASLGACEIPRRDYLNYLNQMRLGR
                LPNNFWVPRCLFSPQE"
     gene       complement(1472. .3193)
                /gene="cydC"
                /locus_tag="S0886"
     CDS        complement(1472. .3193)
                /gene="cydC"
                /locus_tag="S0886"
                /function="transport; Transport of small molecules: Other"
                /note="residues 1 to 573 of 573 are 97.03 pct identical to
                residues 1 to 573 of 573 from Escherichia coli K-12 :
                B0886"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding component of cytochrome-related
                transport"
                /protein_id="AAP16350.1"
                /db_xref="GI:30040619"
                /translation="MRALLPYLALYKRHKWMLSLGIVLAIVTLLASIGLLTLSGWFLS
                ASAVAGVAGLYSFNYMLPAAGVRGAAITRTAGRYFERLVSHDATFRVLQHLRIYTFSK
                LLPLSPAGLARYRQGELLNRVVADVDTLDHLYLRVISPLVGAFVVIMVVTIGLSFLDF
                TLAFTLGSVMLLTLFLMPPLFYRAGKSTGQNLTHLRGQYRQQLTAWLQGQAELTIFGA
                SDRYRTQLENTEIQWLEAQRRQSELTALSQAIMLLIGALAVILMLWMASGGVGGNAQP
                GALIALFVFCALAAFEALAPVTGAFQHLGQVIASAVRISDLTDQKPEVTFPDTQTRVA
                DRVSLTLRDVQFTYPEQSQQALKGISLQVNAGEHIAILGRTGCGKSTLLQLLTRAWDP
                QQGEILLNDSPIASLNEAALRQTISVVPQRVHLFSATLRDNLLLASPGSSDEALSEIL
                RRVGLEKLLEDAGLNSWLGEGGRQLSGGELRRLAIARTLLHDAPLVLLDEPTEGLDAT
                TESQILELLAKMMREKTVLMVTHRLRGLSRFQQIIVMDNGQIIEQGTHAELLARQGRY
                YQFKQGL"
     gene       complement(3194. .4960)
                /gene="cydD"
                /locus_tag="S0887"
     CDS        complement(3194. .4960)
                /gene="cydD"
                /locus_tag="S0887"
                /function="transport; Not classified"
                /note="residues 1 to 588 of 588 are 92.85 pct identical to
                residues 1 to 588 of 588 from Escherichia coli K-12 :
                B0887"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding component of cytochrome-related
                transport"
                /protein_id="AAP16351.1"
                /db_xref="GI:30040620"
                /translation="MNKSRQKELTRWLKQQSVISQRWLNISRLLGFVSGILIIAQAWF
                MARILQHMIMENIPREALLLPFTLLVLTFVLRAWVVWLRERVGYHAGQHIRFAIRRQV
                LDRLQQAGPAWIQGKPAGSWATLVLEQIDDMHDYYARYLPQMALAVSVPLLIVVAIFP
                SNWAAALILLGTAPLIPLFMALVGMGAADANRRNFLALARLSGHFLDRLRGMETLRIF
                GRGEAEIESISRSASEDFRQRTMEVLRLAFLSSGILEFFTSLSIALVAVYFGFSYLGEL
                DFGHYDTGVTLAAGFLALILAPEFFQPLRDLGTFYHAKAQAVGAADSLKTFMETPLAH
                PQRGEAELALTDPLTIEAEDLFITSPEGKTLAGPLNFTLPAGQRAVLVGRSGSGKSSL
                LNALSGFLSYQGSLRINGIELRDLSPESWRKHLSWVGQNPQLPAAKLRDNVLLARPDA
                SEQELQAALDNAWVSEFLPLLPQGVDTPVGDQAARLSVGQAQRVAVARALLNPCSLLL
                LDEPAASLDAHSEQRVMEALNAASLRQTTLMVTHQLEDLADWDVIWVMQDGRIIEQGR
                YAELSVAGGPFATLLAHRQEEI"
     gene       complement(5083. .6048)
                /gene="trxB"
                /locus_tag="S0888"
     CDS        complement(5083. .6048)
                /gene="trxB"
                /locus_tag="S0888"
                /function="enzyme; 2-Deoxyribonucleotide metabolism"
                /note="residues 1 to 321 of 321 are 95.95 pct identical to
                residues 1 to 321 of 321 from Escherichia coli K-12 :
                B0888"
                /codon_start=1
                /transl_table=11
                /product="thioredoxin reductase"
                /protein_id="AAP16352.1"
                /db_xref="GI:30040621"
                /translation="MGTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQL
                TTTTEVENWPGDPNDLTGPLLMARMHEHATKFETEIIFDHINKVDLQNRPFRLNGDNG
                EYTCDALIIATGASARYLGLPSEEAFKGRGVSACATCDGFFYRNQKVAVIGGGNTAVE
                EALYLSNIASEVHLIHRRDGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGV
                TGVRLRDTQNSDNIESLDVAGLFVAIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQ
                TSIPGVFAAGDVMDHIYRQAITSAGTGCMAALDAERYLDGLADAK"
     gene       6592. .7086
                /gene="lrp"
                /locus_tag="S0889"
     CDS        6592. .7086
                /gene="lrp"
                /locus_tag="S0889"
                /function="regulator; Central intermediary metabolism:
                Pool, multipurpose conversions"
                /note="residues 1 to 164 of 164 are 100.00 pct identical
                to residues 1 to 164 of 164 from Escherichia coli K-12 :
                B0889"
                /codon_start=1
                /transl_table=11
                /product="regulator for leucine (or lrp) regulon and
                high-affinity branched-chain amino acid transport system"
                /protein_id="AAP16353.1"
                /db_xref="GI:30040622"
                /translation="MVDSKKRPGKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTP
                CLERVRRLERQGFIQGYTALLNPHYLDASLLVFVEITLNRGAPDVFEQFNTAVQKLEE
                IQECHLVSGDFDYLLKTRVPDMSAYRKLLGETLLRLPGVNDTRTYVVMEEVKQSNRLV
                IKTR"
     gene       7221. .11249
                /gene="ftsK"
                /locus_tag="S0890"
     CDS        7221. .11249
                /gene="ftsK"
                /locus_tag="S0890"
                /function="phenotype; Cell division"
                /note="residues 1 to 737 of 1342 are 68.24 pct identical
                to residues 1 to 737 of 1329 from Escherichia coli K-12 :
                B0890"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein"
                     /protein_id="AAP16354.1"
                     /db_xref="GI:30040623"
                     /translation="MSQEYTEDKEVTLTKLSSGRRLLEALLILIVLFAVWLMAALLSF
                     NPSDPSWSQTAWHEPIHNLGGMPGAWLADTLFFIFGVMAYTIPVIIVGGCWFAWRHQS
                     SDEYIDYFAVSLRIIGVLALILTSCGLAAINADDIWYFASGGVIGSLLSTTLQPLLHS
                     SGGTIALLCVWAAGLTLFTGWSWVTIAEKLGGWILNILTFASNRTRRDDTWVDEDEYE
                     DDEEYEDENHGKQHESRRARILRGALARRKRLAEKFINPMGRQTDAALFSGKRMDDDE
                     EITYTARGVAADPDDVLFSGNRATQPEYDEYDPLLNSAPITEPVAVAAAATTATQSWA
                     APVEPVTQTPPVASVDVPPSQPTVAWQPVPGPQTGEPVIAPAPEGYPQQSQYAQPAVQ
                     YNEPLQQPVQPQQPYYAPAAEQPAQQPYYAPAPEQPVAGNAWQAEEQQSTFAPLSTYQ
                     TEQTYQQPAAQEPLYQQPQPVEQQPVVEPEPVVEETKPARPPLYYFEEVEEKRARERE
                     QLAAWYQPIPEPVKEPEPIKSSLKAPSVAAVPPVEAAAAVSPLASGVKKATLATGAAA
                     TVAAPVFSLANSGGPRPQVKEGIGPQLPRPKRIRVPTRRELASYGIKLPSQRAAEEKA
                     REAQRNQYDSGDQYNDDEIDAMQQDELARQFAQTQQQRYGEQYQHDVPVNAEDADAAA
                     EAELARQFAQTQQQRYSGEQPAGANPFSLDDFEFSPMKALLDDGPHEPLFTPIVEPVQ
                     QPQQPVAPQQQYQQPQQPVAPQQQYQQPQQQVAPQPQYQQPQQPVAPQQQYQQPQQPV
                     APQPQYQQPQQPVAPQPQYQQPQQPVAPQPQDTLLHPLLMRNGDSRPLHKPTTPLPSL
                     DLLTPPPSEVEPVDTFALEQMARLVEARLADFRIKADVVNYSPGPVITRFELNLAPGV
                     KAARISNLSRDLARSLSTVAVRVVEVIPGKPYVGLELPNKKRQTVYLREVLDNAKFRD
```

Query Match 74.7%; Score 340.8; DB 1; Length 292504;
Best Local Similarity 98.0%; Pred. No. 8.3e-88;
Matches 345; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
              | |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     212560 CAGTAATAATAGCGGCCCAAATTCAGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 212619

Qy        165 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212620 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 212679

Qy        225 CGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212680 TGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 212739

Qy        285 TGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGT 344
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||| ||
Db     212740 TGGCTTCGGTAACAGCGCCACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACTGT 212799

Qy        345 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGT 404
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     212800 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCGAACTCCTCCGT 212859

Qy        405 CAACGTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     212860 CAACGTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTATTAA 212911
```

```
RESULT 12
AF237726/c
LOCUS       AF237726                437 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Shigella flexneri insertion sequence IS600, complete sequence.
ACCESSION   AF237726
VERSION     AF237726.1  GI:7582371
KEYWORDS    .
SOURCE      Shigella flexneri
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Sakellaris,H., Hannink,N.K., Rajakumar,K., Bulach,D., Hunt,M.,
            Sasakawa,C. and Adler,B.
  TITLE     Curli loci of Shigella spp
  JOURNAL   Infect. Immun. 68 (6), 3780-3783 (2000)
  MEDLINE   20278182
   PUBMED   10816548
REFERENCE   2  (bases 1 to 437)
  AUTHORS   Sakellaris,H., Hannink,N.K., Rajakumar,K., Bulach,D., Hunt,M.,
            Sasakawa,C. and Adler,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2000) Microbiology, Monash University, Wellington
            Rd., Clayton, Vic 3800, Australia
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Shigella flexneri"
                     /mol_type="genomic DNA"
                     /strain="SBA1100"
                     /serotype="2a"
                     /db_xref="taxon:623"
     source          320. .392
                     /organism="Shigella flexneri"
                     /mol_type="genomic DNA"
                     /strain="SBA1100"
                     /serotype="2a"
                     /db_xref="taxon:623"
                     /insertion_seq="IS600"
                     /note="similar to IS600 deposited in GenBank Accession
                     Number X05952"
     misc_feature    join(1. .319,393. .437)
                     /note="csg locus; similar to Escherichia coli K12 csg
                     locus deposited in GenBank Accession Number X90754"
ORIGIN
```

Query Match 68.3%; Score 311.4; DB 1; Length 437;
Best Local Similarity 98.1%; Pred. No. 1.2e-79;
Matches 315; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy        105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
              | |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        321 CAGTAATAATAGCGGCCCAAATTCAGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 262

Qy        165 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 202

Qy        225 CGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 142

Qy        285 TGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGT 344
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||| ||
Db        141 TGGCTTCGGTAACAGCGCCACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACTGT 82

Qy        345 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGT 404
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         81 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCGAACTCCTCCGT 22

Qy        405 CAACGTGACTCAGGTTGGCTT 425
              |||||||||||||||||||||
Db         21 CAACGTGACTCAGGTTGGCTT 1
```

```
RESULT 13
STAJ2301
LOCUS       STAJ2301                5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
```

```
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
     gene            3344. .3376
                     /gene="csgBA"
     -35_signal      3344. .3349
                     /gene="csgBA"
     -10_signal      3371. .3376
                     /gene="csgBA"
     gene            3469. .3924
                     /gene="csgB"
     CDS             3469. .3924
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="CAA05316.1"
                     /db_xref="GI:2739238"
                     /db_xref="GOA:P55226"
                     /db_xref="SWISS-PROT:P55226"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     3469. .3531
                     /gene="csgB"
     mat_peptide     3532. .3921
                     /gene="csgB"
                     /product="CsgB protein"
                     /function="nucleator"
     gene            3966. .4421
                     /gene="csgA"
     CDS             3966. .4421
                     /gene="csgA"
                     /function="fiber subunit"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="CAA05317.1"
                     /db_xref="GI:2739239"
                     /db_xref="GOA:P55225"
                     /db_xref="SWISS-PROT:P55225"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     3966. .4025
                     /gene="csgA"
     mat_peptide     4026. .4418
                     /gene="csgA"
                     /product="CsgA protein"
                     /function="fiber subunit"
     gene            4483. .4809
                     /gene="csgC"
     CDS             4483. .4809
                     /gene="csgC"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="CsgC protein"
                     /protein_id="CAA05318.1"
                     /db_xref="GI:2739240"
                     /db_xref="SWISS-PROT:P55227"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
     sig_peptide     4483. .4533
                     /gene="csgC"
     mat_peptide     4534. .4806
                     /gene="csgC"
                     /product="CsgC"
     gene            4940. .5103
                     /gene="ORF103"
     CDS             4940. .>5103
                     /gene="ORF103"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA05319.1"
                     /db_xref="GI:2739241"
                     /db_xref="SWISS-PROT:O54295"
                     /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                     CDISTHAENI"
ORIGIN

  Query Match           56.8%;  Score 259.2;  DB 1;  Length 5103;
  Best Local Similarity  73.0%;  Pred. No. 2.9e-64;
  Matches  333;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  ||||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              | |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| ||||  | || || ||||| ||   ||| ||| | ||||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 14
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .22411
                     /organism="Salmonella typhimurium LT2"
                     /mol_type="genomic DNA"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            434. .1308
                     /gene="phoH"
                     /note="synonym: STM1126"
     RBS             434. .439
                     /gene="phoH"
                     /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS             454. .1308
                     /gene="phoH"
                     /note="similar to E. coli PhoB-dependent, ATP-binding pho
                     regulon component; may be helicase; induced by P
                     starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                     aa), 92% identity in aa 71 - 354"
                     /codon_start=1
                     /transl_table=11
                     /product="PhoB-dependent ATP-binding pho regulon
                     component"
                     /protein_id="AAL20057.1"
                     /db_xref="GI:16419642"
                     /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                     RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                     GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                     YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
```

LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
QRTLHAYS"
gene complement(1414. .2303)
/gene="STM1127"
CDS complement(1414. .2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
identity in aa 15 - 264"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
QARESKLRTMNALDMTK"
RBS complement(2298. .2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"
gene complement(2580. .4088)
/gene="STM1128"
CDS complement(2580. .4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter
(AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
identity in aa 7 - 478"
/codon_start=1
/transl_table=11
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSPGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
DKFPPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
NAKLVAVIPVFPFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
RYATDLNFFPYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
RBS complement(4083. .4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"
gene complement(4413. .5099)
/gene="STM1129"
CDS complement(4413. .5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
WFSHAVKR"
RBS complement(5094. .5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"
gene 5599. .6759
/gene="STM1130"
CDS 5599. .6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
identity in aa 40 - 403"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKPYSHEGINKKWRDEVYGLINGH
WQYMGKMKQPLGYGVSVSYGDEVPLIGGENAKGKPVSSVTSFTMRDGNLLIK"
gene 6793. .7497
/gene="STM1131"
RBS 6793. .6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"
CDS 6805. .7497
/gene="STM1131"
/note="similar to E. coli orf, hypothetical protein
(AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
identity in aa 4 - 239"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein"
/protein_id="AAL20062.1"
/db_xref="GI:16419647"
/translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
QTRYRVGLQYHF"
gene 7768. .9060
/gene="STM1132"
RBS 7768. .7773
/gene="STM1132"
/note="putative RBS for STM1132; RegulonDB:STMS1H001403"
CDS 7780. .9060
/gene="STM1132"
/note="similar to E. coli putative transport protein
(AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
identity in aa 25 - 418"
/codon_start=1
/transl_table=11
/product="putative sugar transport protein"

Query Match 56.8%; Score 259.2; DB 1; Length 22411;
Best Local Similarity 73.0%; Pred. No. 3.6e-64;
Matches 333; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  ||||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              | |||||||||| |||||   ||| |||||||||    ||     ||||| || || |||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| ||||  | || || ||||| ||   ||| ||| | ||||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
```

```
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224
```

RESULT 15
AL627269

```
LOCUS       AL627269              254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
```

```
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
misc_feature         complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
gene                 5924. .8032
                     /gene="STY1082"
CDS                  5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
misc_feature         5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
misc_feature         6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
gene                 8045. .9952
                     /gene="STY1083"
CDS                  8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap

Query Match           56.5%;  Score 257.6;  DB 1;  Length 254050;
Best Local Similarity  72.8%;  Pred. No. 1.5e-63;
Matches  332;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || | |||||||||||||| ||     || ||||| |||| |   |||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  ||||   | ||| ||||| |||||||||  || |||| |||| ||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| ||||  | || || ||||| ||   ||| ||| | ||||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || |||||||||  ||||| || || || || ||    || | ||| |||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  ||| ||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361

Search completed: March 18, 2004, 15:02:31
Job time : 2313.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	3	AAC64619	Aac64619 Escherich
2	456	100.0	456	9	ACF36153	Acf36153 E. coli C
3	437.2	95.9	646	2	AAQ62647	Aaq62647 Fibronect
4	355.4	77.9	369	2	AAQ62646	Aaq62646 FNB curli
5	254.4	55.8	456	2	AAQ87467	Aaq87467 AgfA sequ
6	254.4	55.8	456	2	AAT74142	Aat74142 Salmonell
7	254.4	55.8	456	3	AAC64617	Aac64617 Salmonell
8	238.4	52.3	456	3	AAC64628	Aac64628 AgfA::PT3
9	236.8	51.9	456	3	AAC64626	Aac64626 AgfA::PT3
10	230.4	50.5	456	3	AAC64625	Aac64625 AgfA::PT3
11	230.4	50.5	456	3	AAC64627	Aac64627 AgfA::PT3
12	228.8	50.2	456	3	AAC64630	Aac64630 AgfA::PT3
13	227.2	49.8	456	3	AAC64631	Aac64631 AgfA::PT3
14	225.6	49.5	456	3	AAC64623	Aac64623 AgfA::PT3
15	222.4	48.8	456	3	AAC64624	Aac64624 AgfA::PT3
16	220.8	48.4	456	3	AAC64629	Aac64629 AgfA::PT3
17	216	47.4	456	3	AAC64622	Aac64622 AgfA::PT3
18	173.4	38.0	361	2	AAQ73066	Aaq73066 AgfA sequ
19	173.4	38.0	361	2	AAT74141	Aat74141 Salmonell
20	100	21.9	100	7	ACD68807	Acd68807 E. coli K
21	100	21.9	100	7	ACD68808	Acd68808 E. coli K
22	100	21.9	100	7	ACD68809	Acd68809 E. coli K
23	78	17.1	78	9	ACF36151	Acf36151 E. coli C
24	72	15.8	72	9	ACF36148	Acf36148 E. coli C
25	66	14.5	66	9	ACF36150	Acf36150 E. coli C
26	43.2	9.5	1083	5	AAS76745	Aas76745 DNA encod
27	40.2	8.8	4018	2	AAQ63879	Aaq63879 Polyhydro
28	39	8.6	1390	9	ADC86984	Adc86984 Human GPC
c 29	39	8.6	2730	5	AAS87782	Aas87782 DNA encod
c 30	37.4	8.2	1972	9	ADC87176	Adc87176 Human GPC
31	37	8.1	2580	5	AAH66822	Aah66822 C glutami
c 32	37	8.1	349980	5	AAH68529	Aah68529 C glutami
33	36	7.9	36	9	ACF36152	Acf36152 E. coli V
34	36	7.9	36	9	ACF36149	Acf36149 E. coli C
c 35	35.8	7.9	3984	4	ABL27701	Abl27701 Drosophil
36	35.8	7.9	12775	4	ABL27700	Abl27700 Drosophil
37	34.8	7.6	1200	5	AAS92828	Aas92828 DNA encod
38	34.6	7.6	1633	9	ADC87022	Adc87022 Human GPC
c 39	34.6	7.6	3893	6	ABK52154	Abk52154 Haemophil
40	34.6	7.6	155074	6	ABN85735	Abn85735 Human gen
c 41	34.2	7.5	1543	5	AAS92206	Aas92206 DNA encod
42	34.2	7.5	2000	7	ADA71938	Ada71938 Rice gene
43	34.2	7.5	3532	9	ADC87116	Adc87116 Human GPC
44	34.2	7.5	38918	9	ADC87242	Adc87242 Human GPC
c 45	34	7.5	570	4	ABA36458	Aba36458 Probe #14

ALIGNMENTS

RESULT 1
AAC64619
ID AAC64619 standard; DNA; 456 BP.
XX
AC AAC64619;
XX
DT 26-FEB-2001 (first entry)
XX
DE Escherichia coli CsgA DNA sequence SEQ ID NO:3.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Escherichia coli.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36343.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 122 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 5.1e-134;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

RESULT 2
ACF36153
ID ACF36153 standard; DNA; 456 BP.
XX
AC ACF36153;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein encoding DNA.
XX
KW Plasma protein; immune response; antibacterial; vaccine; gene therapy;
KW gene; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1. .456
FT /*tag= a
FT /note= "15 kDa protein"
XX
PN WO2003064446-A2.
XX
PD 07-AUG-2003.
XX
PF 30-JAN-2003; 2003WO-EP000943.
XX
PR 31-JAN-2002; 2002GB-00002275.
XX
PA (HANS-) HANSA MEDICAL RES AB.
XX
PI Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
DR WPI; 2003-646136/61.
DR P-PSDB; ABR82651.
XX
PT New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
PS Disclosure; Page 41-42; 42pp; English.
XX
CC The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein encoding DNA
XX
SQ Sequence 456 BP; 113 A; 110 C; 122 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 9; Length 456;
Best Local Similarity 100.0%; Pred. No. 5.1e-134;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

RESULT 3
AAQ62647
ID AAQ62647 standard; DNA; 646 BP.
XX
AC AAQ62647;
XX
DT 25-MAR-2003 (revised)
DT 18-OCT-1994 (first entry)
XX
DE Fibronectin binding protein fnbA gene.
XX
KW Fibronectin binding protein; FBP; curlin; curli pili; pilus;
KW immunization; vaccine; fnbA gene; ss.
XX
OS Escherichia coli.
XX
PN WO9410330-A1.
XX
PD 11-MAY-1994.
XX
PF 03-NOV-1993; 93WO-US010547.
XX
PR 03-NOV-1992; 92US-00970846.
XX
PA (REAT/) REA T S.
XX
PI Normark S, Olsen A;
XX
DR WPI; 1994-167484/20.
DR P-PSDB; AAR52664.
XX
PT New fibronectin binding protein curlin from E. coli - used for the
PT treatment of infections caused by E. coli and for immunisation against E.
PT coli infection.
XX
PS Disclosure; Fig 2; 33pp; English.
XX
CC The fnbA gene of E. coli cattle fecal isolate AO12 includes DNA encoding
CC a new fibronectin binding protein (FBP), curlin, which is produced as
CC curli pili when expressed in E. coli (pFnb20) DSM 4585. The fnbA gene
CC sequence from pFnb20 is given in AAQ62647, and includes the sequence
CC encoding the 17 kDa curlin subunit (amino acids 21-142 of AAR52664).
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 646 BP; 168 A; 146 C; 161 G; 171 T; 0 U; 0 Other;

Query Match 95.9%; Score 437.2; DB 2; Length 646;
Best Local Similarity 98.9%; Pred. No. 5.4e-128;
Matches 451; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGAGCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTC--TCGTCAACGTGACTCAGGTT 500

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        501 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 536
```

RESULT 4
AAQ62646
ID AAQ62646 standard; DNA; 369 BP.
XX
AC AAQ62646;
XX
DT 25-MAR-2003 (revised)
DT 18-OCT-1994 (first entry)
XX
DE FNB curlin.
XX
KW Fibronectin binding protein; FBP; curlin; curli pili; pilus;
KW immunization; vaccine; ss.
XX
OS Escherichia coli.
XX
PN WO9410330-A1.
XX
PD 11-MAY-1994.
XX
PF 03-NOV-1993; 93WO-US010547.
XX
PR 03-NOV-1992; 92US-00970846.
XX
PA (REAT/) REA T S.
XX
PI Normark S, Olsen A;
XX
DR WPI; 1994-167484/20.
DR P-PSDB; AAR52663.
XX
PT New fibronectin binding protein curlin from E. coli - used for the
PT treatment of infections caused by E. coli and for immunisation against E.
PT coli infection.
XX
PS Claim 8; Page 22; 33pp; English.
XX
CC The fnbA gene of E. coli cattle fecal isolate AO12 includes DNA encoding
CC a new fibronectin binding protein, curlin, which is produced as curli
CC pili when expressed in E. coli (pFnb20) DSM 4585. (Updated on 25-MAR-2003
CC to correct PN field.)
XX
SQ Sequence 369 BP; 86 A; 87 C; 105 G; 91 T; 0 U; 0 Other;

Query Match 77.9%; Score 355.4; DB 2; Length 369;
Best Local Similarity 99.2%; Pred. No. 3.5e-102;
Matches 368; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

```
Qy         60 AGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGG 119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGG 60

Qy        120 CCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCA 120

Qy        180 AACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGA 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGA 180

Qy        240 TGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAG 299
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAG 240

Qy        300 CGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGG 359
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGG 300

Qy        360 TGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGT 419
                |||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        301 TGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTC--TCGTCAACGTGACTCAGGT 358

Qy        420 TGGCTTTGGTA 430
                ||||||||||||
Db        359 TGGCTTTGGTA 369


RESULT 5
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag=  a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             55.8%;  Score 254.4;  DB 2;  Length 456;
  Best Local Similarity   72.4%;  Pred. No. 3.9e-70;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
               || || ||||| || |  ||||||||||||||| ||    || |||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
               ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
               |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
               || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
               || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
               || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAT74142
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
```

CC family Enterobacteria. It can also be used to provide proteins and
CC antibodies which can be used for assays. The nucleic acid sequence can be
CC used to provide probes or primers which can specifically hybridise to
CC nucleic acid molecules from greater than 99% of Salmonella strains that
CC are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC correct PF field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 55.8%; Score 254.4; DB 2; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.9e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 7
AAC64617
ID AAC64617 standard; DNA; 456 BP.
XX
AC AAC64617;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36341.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 55.8%; Score 254.4; DB 3; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.9e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 8
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match          52.3%;  Score 238.4;  DB 3;  Length 456;
  Best Local Similarity  70.2%;  Pred. No. 4.5e-65;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360
Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy        421 GGCTTTGGTAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 9
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match             51.9%;  Score 236.8;  DB 3;  Length 456;
  Best Local Similarity   70.0%;  Pred. No. 1.5e-64;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| | |||   |    ||||||    ||    |  |  |   ||| ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             50.5%;  Score 230.4;  DB 3;  Length 456;
  Best Local Similarity   69.1%;  Pred. No. 1.6e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
```

```
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match           50.5%;  Score 230.4;  DB 3;  Length 456;
  Best Local Similarity  69.1%;  Pred. No. 1.6e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64630
ID   AAC64630 standard; DNA; 456 BP.
XX
AC   AAC64630;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

```
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36354.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match           50.2%;  Score 228.8;  DB 3;  Length 456;
  Best Local Similarity  68.9%;  Pred. No. 5e-62;
  Matches  314;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||||  || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
                || ||||                 | | ||   |    || ||  ||| | || |
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
```

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 49.8%; Score 227.2; DB 3; Length 456;
Best Local Similarity 68.6%; Pred. No. 1.6e-61;
Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
Qy    1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
          ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db    1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy   61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
          || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db   61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy  121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
          ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db  121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy  181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
          |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db  181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy  241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
          || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db  241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy  301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
          || ||   |||||||  |     |   |  |          |   |||   || |||
Db  301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy  361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
            ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db  361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy  421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
            || ||||| |||||||| || ||| | ||||| |||
Db  421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 14
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 49.5%; Score 225.6; DB 3; Length 456;
Best Local Similarity 68.4%; Pred. No. 5.2e-61;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```
Qy    1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
          |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db    1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy   61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
          || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db   61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy  121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
          ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db  121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy  181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
          |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db  181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy  241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
          || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db  241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy  301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
          || ||   | || |||||||||| ||||| || || || || ||    || | | |
Db  301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy  361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||  |||  |||         |  |  |   ||| |  |||  |||||||
Db  361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy  421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
            || ||||| |||||||| || ||| | ||||| |||
Db  421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 15
AAC64624
ID AAC64624 standard; DNA; 456.BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 48.8%; Score 222.4; DB 3; Length 456;
Best Local Similarity 68.0%; Pred. No. 5.3e-60;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || ||  | |||| |  | | | |  | ||  |  ||  |  ||   ||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||| || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || ||||| | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

Search completed: March 18, 2004, 09:11:49
Job time : 290.405 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	254.4	55.8	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	173.4	38.0	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	35.6	7.8	1788	4	US-09-252-991A-4632	Sequence 4632, Ap
	4	35.6	7.8	2250	4	US-09-252-991A-4181	Sequence 4181, Ap
c	5	35.6	7.8	3102	4	US-09-252-991A-4429	Sequence 4429, Ap
c	6	34	7.5	552	4	US-09-252-991A-2024	Sequence 2024, Ap
	7	34	7.5	669	4	US-09-252-991A-1852	Sequence 1852, Ap
	8	34	7.5	1185	4	US-09-252-991A-1931	Sequence 1931, Ap
	9	33.4	7.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	10	33.4	7.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	11	33	7.2	1830121	4	US-09-557-884-1	Sequence 1, Appli
	12	33	7.2	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c	13	32	7.0	2993	2	US-08-225-488-6	Sequence 6, Appli
	14	31.8	7.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	15	31.4	6.9	1317	4	US-09-252-991A-10989	Sequence 10989, A
	16	31.2	6.8	1830	4	US-09-667-135-29	Sequence 29, Appl
	17	31	6.8	1422	4	US-09-328-352-586	Sequence 586, App
	18	30.6	6.7	435	3	US-08-569-147-75	Sequence 75, Appl
	19	30.6	6.7	1923	4	US-09-056-556-190	Sequence 190, App
	20	30.6	6.7	1923	4	US-09-072-596-185	Sequence 185, App
	21	30.6	6.7	1923	4	US-09-072-967-190	Sequence 190, App
	22	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	23	30	6.6	6122	1	US-08-403-545-1	Sequence 1, Appli
	24	30	6.6	6122	3	US-08-404-381-1	Sequence 1, Appli
	25	30	6.6	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	26	29.8	6.5	771	4	US-09-489-039A-915	Sequence 915, App
	27	29.8	6.5	1128	4	US-09-328-352-3993	Sequence 3993, Ap
	28	29.8	6.5	2652	4	US-09-489-039A-5584	Sequence 5584, Ap
c	29	29.8	6.5	87563	4	US-09-453-702B-57	Sequence 57, Appl
	30	29.2	6.4	14654	4	US-08-961-527-106	Sequence 106, App
c	31	28.8	6.3	429	4	US-09-252-991A-8302	Sequence 8302, Ap
c	32	28.8	6.3	912	2	US-08-993-228-7	Sequence 7, Appli
	33	28.8	6.3	1023	4	US-09-328-352-3184	Sequence 3184, Ap
	34	28.8	6.3	1395	4	US-09-252-991A-8312	Sequence 8312, Ap
	35	28.8	6.3	9507	4	US-09-489-039A-6896	Sequence 6896, Ap
	36	28.8	6.3	12720	1	US-08-403-866-11	Sequence 11, Appl
	37	28.6	6.3	217	4	US-09-313-294A-5247	Sequence 5247, Ap
	38	28.6	6.3	693	4	US-09-328-352-3008	Sequence 3008, Ap
	39	28.6	6.3	1507	4	US-09-453-323-1	Sequence 1, Appli
	40	28.6	6.3	1873	3	US-09-461-474-9	Sequence 9, Appli
	41	28.4	6.2	1380	4	US-09-489-039A-2442	Sequence 2442, Ap
c	42	28.4	6.2	62909	4	US-09-596-002-32	Sequence 32, Appl
	43	28.2	6.2	298	4	US-09-313-294A-4713	Sequence 4713, Ap
	44	28.2	6.2	346	3	US-09-060-756-317	Sequence 317, App
	45	28.2	6.2	346	4	US-09-670-314-317	Sequence 317, App

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 55.8%; Score 254.4; DB 1; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.3e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 360

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
;
US-08-233-788A-56

Query Match 38.0%; Score 173.4; DB 1; Length 361;
Best Local Similarity 69.9%; Pred. No. 7.3e-45;
Matches 234; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```
Qy         64 GTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCA 123
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 AATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACT 183
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTT 243
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCT 303
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGC 363
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGGTGCTGCAGTTGACCAGACTGCATCTAACTC 398
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 3
US-09-252-991A-4632/c
; Sequence 4632, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4632
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4632

Query Match 7.8%; Score 35.6; DB 4; Length 1788;
Best Local Similarity 64.6%; Pred. No. 0.23;
Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy         30 AGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGG 89
Db       1391 AGCGGGTGTATTCGCCGGTGCCGGACTGCGCGGTGGCGATCGGCAGGAAGCCGGCGGCGG 1332

Qy         90 TAACCACGGTGGTGGCGGTAAT 111
Db       1331 TGATCAGGGTGCCGGTGAGCAT 1310
```

```
RESULT 4
US-09-252-991A-4181
; Sequence 4181, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 4181
;   LENGTH: 2250
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4181

  Query Match             7.8%;  Score 35.6;  DB 4;  Length 2250;
  Best Local Similarity   64.6%;  Pred. No. 0.26;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         30 AGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGG 89
              |||    |||||| |||||  ||  |||   ||||  | ||  ||| | | |||||||||
Db        339 AGCGGGTGTATTCGCCGGTGCCGGACTGCGCGGTGGCGATCGGCAGGAAGCCGGCGGCGG 398

Qy         90 TAACCACGGTGGTGGCGGTAAT 111
              | | || ||||  || |   ||
Db        399 TGATCAGGGTGCCGGTGAGCAT 420


RESULT 5
US-09-252-991A-4429/c
; Sequence 4429, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 4429
;   LENGTH: 3102
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4429

  Query Match             7.8%;  Score 35.6;  DB 4;  Length 3102;
  Best Local Similarity   64.6%;  Pred. No. 0.3;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         30 AGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGG 89
              |||    |||||| |||||  ||  |||   ||||  | ||  ||| | | |||||||||
Db       1411 AGCGGGTGTATTCGCCGGTGCCGGACTGCGCGGTGGCGATCGGCAGGAAGCCGGCGGCGG 1352

Qy         90 TAACCACGGTGGTGGCGGTAAT 111
              | | || ||||  || |   ||
Db       1351 TGATCAGGGTGCCGGTGAGCAT 1330


RESULT 6
US-09-252-991A-2024/c
; Sequence 2024, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2024
;   LENGTH: 552
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2024

  Query Match             7.5%;  Score 34;  DB 4;  Length 552;
  Best Local Similarity   56.1%;  Pred. No. 0.44;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        147 GTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGAC 206
              || ||| |||| |  |  ||  | || || ||   | |||| ||   |  ||  |  |
Db        119 GTGCGGCGGCGCTGGCGGTGTTCGTGTTCGGCGTCCCGATGCGCGGCTCCTTGCTGTCCC 60

Qy        207 TATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGA 260
              | || | |      ||| | | | |||||||  |  | |||    || || | |||
Db         59 TGTTGCTCTTGTCGGCCGTGTTCATGGTGCCGGCGCTGGGTCAGGGTCTGCTGA 6


RESULT 7
US-09-252-991A-1852
; Sequence 1852, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1852
;   LENGTH: 669
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1852

  Query Match             7.5%;  Score 34;  DB 4;  Length 669;
  Best Local Similarity   56.1%;  Pred. No. 0.48;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        147 GTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGAC 206
              || ||| |||| |  |  ||  | || || ||   | ||||| ||   |  ||  |  |
Db        462 GTGCGGCGGCGCTGGCGGTGTTCGTGTTCGGCGTCCCGATGCGCGGCTCCTTGCTGTCCC 521

Qy        207 TATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGA 260
              | || | |      ||| | | | |||||||  |  | |||    || || | |||
Db        522 TGTTGCTCTTGTCGGCCGTGTTCATGGTGCCGGCGCTGGGTCAGGGTCTGCTGA 575


RESULT 8
US-09-252-991A-1931
; Sequence 1931, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
```

```
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1931
;   LENGTH: 1185
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1931

  Query Match             7.5%;  Score 34;  DB 4;  Length 1185;
  Best Local Similarity   56.1%;  Pred. No. 0.62;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        147 GTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGAC 206
              || ||| |||| |  |  ||  | || || ||  | ||||| ||   |  ||   | |
Db        782 GTGCGGCGGCGCTGGCGGTGTTCGTGTTCGGCGTCCCGATGCGCGGCTCCTTGCTGTCCC 841

Qy        207 TATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGA 260
              | || | |      ||| | | | |||||||| |  | |||    || || | |||
Db        842 TGTTGCTCTTGTCGGCCGTGTTCATGGTGCCGGCGCTGGGTCAGGGTCTGCTGA 895


RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             7.3%;  Score 33.4;  DB 3;  Length 4403765;
  Best Local Similarity   72.9%;  Pred. No. 30;
  Matches   43;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        206 CTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGC 264
              || |  | ||||| | ||||| || |||||| |||||||| |||| || |  |||  ||
Db     371246 CTGTCGCGCAGCAGGCCGGCGATACTGGTGCTGATGTTGGCCAGGCCTGAATTGAAGGC 371304


RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             7.3%;  Score 33.4;  DB 3;  Length 4411529;
  Best Local Similarity   72.9%;  Pred. No. 30;
  Matches   43;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        206 CTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGC 264
              || |  | ||||| | ||||| || |||||| |||||||| |||| || |  |||  ||
Db     371189 CTGTCGCGCAGCAGGCCGGCGATACTGGTGCTGATGTTGGCCAGGCCTGAATTGAAGGC 371247


RESULT 11
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;  GENERAL INFORMATION:
;       APPLICANT: Fleischmann et al.
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/557,884
;            FILING DATE: 25-Apr-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/476,102
;            FILING DATE: JUN-5-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Michelle S. Marks
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB186P3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-309-8504
;            TELEFAX: 301-309-8439
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             7.2%;  Score 33;  DB 4;  Length 1830121;
  Best Local Similarity   55.8%;  Pred. No. 30;
  Matches   63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
```

```
                | | || |     || | ||||  | |    | |   || |||||  | | | |||| |
Db     740925 CGATTTTTCGCAAGATCGTTTAAATGAATATCGCCTACAATTCTGCCTTGGTTATGCACA 740984

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGT 234
              ||||| | ||   |   ||   | |||||    | ||| ||||| ||| |||
Db     740985 TTGATGTCTGTTTTTAAAACCAAATTATTATTTTGAATGTCGGCGTTAAGGGT 741037


RESULT 12
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;  GENERAL INFORMATION:
;       APPLICANT: Robert D. Fleischmann
;                  Mark D. Adams
;                  Owen White
;                  Hamilton O. Smith
;                  J. Craig Venter
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville,
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/643,990A
;            FILING DATE: 23-Aug-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/487,429
;            FILING DATE: 1995-06-07
;            APPLICATION NUMBER: 08/426,787
;            FILING DATE: 1995-04-21
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Kenley K. Hoover
;            REGISTRATION NUMBER: 40,302
;            REFERENCE/DOCKET NUMBER: PB186P1C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-610-5790
;            TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match             7.2%;  Score 33;  DB 4;  Length 1830121;
  Best Local Similarity   55.8%;  Pred. No. 30;
  Matches   63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
              | | || |     || | ||||  | |    | |   || |||||  | | | |||| |
Db     740925 CGATTTTTCGCAAGATCGTTTAAATGAATATCGCCTACAATTCTGCCTTGGTTATGCACA 740984

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGT 234
              ||||| | ||   |   ||   | |||||    | ||| ||||| ||| |||
Db     740985 TTGATGTCTGTTTTTAAAACCAAATTATTATTTTGAATGTCGGCGTTAAGGGT 741037


RESULT 13
US-08-225-488-6/c
; Sequence 6, Application US/08225488
; Patent No. 5846802
;  GENERAL INFORMATION:
;    APPLICANT:  Buxton, Frank
;    APPLICANT:  Hinnen, Albert
;    APPLICANT:  Visser, Jacob
;    TITLE OF INVENTION:  No. 5846802el Fungal Protease
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/225,488
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/047,214
;      FILING DATE:  13-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  4-19055/A/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2993 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;      STRAIN:  N400
;    IMMEDIATE SOURCE:
;      CLONE:  pTZPEPD
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..829
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(830..1153, 1205..1649, 1697..1785, 1841
;      LOCATION:  ..2233)
;      OTHER INFORMATION:  /note= "subtilisin-type protease
;      OTHER INFORMATION:  PEPD of Aspergillus niger; product of gene pepD"
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1154..1204
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1650..1696
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1786..1840
US-08-225-488-6

  Query Match             7.0%;  Score 32;  DB 2;  Length 2993;
  Best Local Similarity   46.4%;  Pred. No. 3.9;
  Matches  104;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;
```

```
Qy          65 TTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAA 124
               ||||   |||  |    || |        ||||  ||||| | | | |    |  ||| |
Db        2181 TTGTAGGCCAGAAGATTGGGGCTACCCGCCACATTGGTGACAGCATTCCGCGTAGCCAAC 2122

Qy         125 ATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTG 184
                 |  ||||||      || |     | | ||  | ||      | |  || |  ||||  |
Db        2121 CTCTTGAGCTCGGTCGTTGCAGCCGCTGGGGTAGCAAGGTCCCGCAAGCCCATCAAATAG 2062

Qy         185 ATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTG 244
               | |  |  | || | || |||    || |||       | | |||  || |||  | ||
Db        2061 AGGATCAAACCTGTCACATGAGGTGTAGCCATGGACGTGCCGGAGATCGTGTTGGTGGCC 2002

Qy         245 GTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGC 288
               |     |  | | |    ||  ||   |  ||   |   || |||
Db        2001 GAGTTCGAGCCGGTCCATGCAGAAAGTACTTGCTCTCCCGGGGC 1958
```

RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 7.0%; Score 31.8; DB 3; Length 4403765;
Best Local Similarity 46.0%; Pred. No. 72;
Matches 108; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy          82 GGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATT 141
               ||||  ||     ||| |||||  |||   |||  ||||          |||       |
Db    4029096 GGCGTTGGGGGTGACGCTGGTGACGGGGGTAACGGCGGGATCGGCGGTGACGGCGGCGCC 4029037

Qy         142 TACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGAC 201
                 |  |||| ||||    |   ||    |  |||       |   || ||| ||| |
Db    4029036 GGCGGAGACGGCGGCGCGGGCGGCGCGGGCGGCCTGTTCGGCGTCTCCGGTAGCTCCGGG 4028977

Qy         202 TTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGAC 261
               ||     | ||     | |||| |||||||| |  |  | |||| ||  |   |  | |
Db    4028976 CTGGGCGGTGCCGCCGGTAGCGGGGGTAATGGCGGGGGCGGTGGTGAGCCCGGTGTGGCC 4028917

Qy         262 AGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGT 316
               ||||    |  |   |    |||  ||  |  ||  |  || |  ||||
Db    4028916 GGCTCCCCGGGCGTTGGGCCTGCGGGCCGTGGCGGCGACGGCAACCTCGGCCAGT 4028862
```

RESULT 15
US-09-252-991A-10989/c
; Sequence 10989, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10989
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10989

Query Match 6.9%; Score 31.4; DB 4; Length 1317;
Best Local Similarity 52.7%; Pred. No. 4.2;
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy          64 GTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCA 123
               |||||  |  ||||  |  |||  |     |||   |   ||  ||  |  |||  |          |||
Db        513 GTTGTCCTTCAGCAGGGCCGGCAGGGTGTCGTGGGAGGGGACGGCGACGTCGATCGCCTC 454

Qy         124 AATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACT 183
                  |||  ||  |   |  ||  ||||||  |    ||||||   |     |  |  ||  |  |
Db        453 GCCACTGCGCAGCGCCTTGTCCAGTTCCTCGGCGGTGGCGAAGGTGTGGTACTCGACGCG 394

Qy         184 GATGCCCGT 192
               ||||||  ||
Db        393 GATGCCGGT 385
```

Search completed: March 18, 2004, 20:20:03
Job time : 76.4971 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

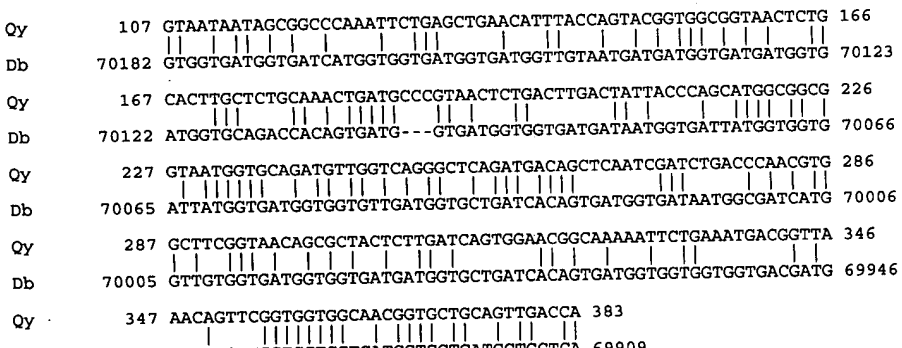
	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	42.6	9.3	135259	12	US-10-240-425-1585	Sequence 1585, Ap
	2	39	8.6	1390	14	US-10-017-161-1781	Sequence 1781, Ap
	3	39	8.6	1390	15	US-10-292-798-1437	Sequence 1437, Ap
	4	38.2	8.4	625	14	US-10-029-386-24675	Sequence 24675, A
c	5	37.4	8.2	1972	14	US-10-017-161-1981	Sequence 1981, Ap
c	6	37.4	8.2	1972	15	US-10-292-798-1629	Sequence 1629, Ap
	7	37	8.1	2580	9	US-09-738-626-1857	Sequence 1857, Ap
c	8	37	8.1	3309400	9	US-09-738-626-1	Sequence 1, Appli
c	9	35.8	7.9	1493	14	US-10-029-386-25133	Sequence 25133, A
c	10	35.6	7.8	992	12	US-10-425-114-31747	Sequence 31747, A
	11	35.6	7.8	2122	15	US-10-108-260A-1242	Sequence 1242, Ap
c	12	35.6	7.8	3078	12	US-10-389-647-226	Sequence 226, App
	13	35.6	7.8	4389	15	US-10-369-493-46798	Sequence 46798, A
c	14	35	7.7	785	14	US-10-029-386-22627	Sequence 22627, A
	15	34.6	7.6	532	14	US-10-029-386-7202	Sequence 7202, Ap
	16	34.6	7.6	1633	14	US-10-017-161-1819	Sequence 1819, Ap
	17	34.6	7.6	1633	15	US-10-292-798-1475	Sequence 1475, Ap
	18	34.6	7.6	155074	13	US-10-026-188-6	Sequence 6, Appli
c	19	34.2	7.5	528	14	US-10-029-386-5350	Sequence 5350, Ap
c	20	34.2	7.5	666	14	US-10-029-386-25948	Sequence 25948, A
	21	34.2	7.5	1692	14	US-10-355-430-33	Sequence 33, Appl
	22	34.2	7.5	3532	14	US-10-017-161-1913	Sequence 1913, Ap
	23	34.2	7.5	3532	15	US-10-292-798-1569	Sequence 1569, Ap
c	24	34.2	7.5	12981	15	US-10-369-493-33865	Sequence 33865, A
	25	34.2	7.5	38918	14	US-10-017-161-2049	Sequence 2049, Ap
	26	34.2	7.5	38918	15	US-10-292-798-1695	Sequence 1695, Ap
c	27	34	7.5	570	9	US-09-864-761-21778	Sequence 21778, A
c	28	34	7.5	1976	9	US-09-864-761-5052	Sequence 5052, Ap
	29	33.8	7.4	765	15	US-10-369-493-44503	Sequence 44503, A
c	30	33.6	7.4	652	15	US-10-260-238-4681	Sequence 4681, Ap
c	31	33.6	7.4	1197	14	US-10-156-761-4579	Sequence 4579, Ap
c	32	33.6	7.4	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	33	33.4	7.3	5163	12	US-10-282-122A-26364	Sequence 26364, A
c	34	33.4	7.3	6615	12	US-10-282-122A-28180	Sequence 28180, A
c	35	33.2	7.3	832	12	US-10-282-122A-19206	Sequence 19206, A
	36	33.2	7.3	1377	14	US-10-238-075-1261	Sequence 1261, Ap
c	37	33.2	7.3	11475	14	US-10-238-075-1250	Sequence 1250, Ap
c	38	33	7.2	379	14	US-10-029-386-19106	Sequence 19106, A
c	39	33	7.2	569	14	US-10-029-386-3710	Sequence 3710, Ap
c	40	33	7.2	652	14	US-10-029-386-22783	Sequence 22783, A
	41	33	7.2	867	12	US-10-282-122A-19705	Sequence 19705, A
	42	33	7.2	1959	9	US-09-864-761-4012	Sequence 4012, Ap
c	43	33	7.2	1973	9	US-09-864-761-3471	Sequence 3471, Ap
c	44	33	7.2	3897	12	US-10-282-122A-22081	Sequence 22081, A
c	45	33	7.2	9590	12	US-10-282-122A-33351	Sequence 33351, A

ALIGNMENTS

RESULT 1
US-10-240-425-1585/c
; Sequence 1585, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1585
; LENGTH: 135259
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 Z83838
US-10-240-425-1585

Query Match 9.3%; Score 42.6; DB 12; Length 135259;
Best Local Similarity 47.5%; Pred. No. 0.011;
Matches 160; Conservative 0; Mismatches 174; Indels 3; Gaps 1;

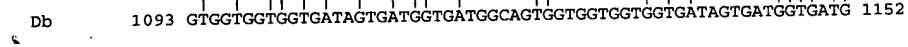
```
Qy         47 GTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCG 106
              ||| | |   | | | || || | |     | |  || || | | | ||| ||| ||| |
Db      70242 GTAATGGTGGTGGTGATGGTGATGGTGGTAATGATGGTGGTGCTGATCACAGTGATGGTG 70183
```



RESULT 2
US-10-017-161-1781
; Sequence 1781, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1781
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1390)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1190)
US-10-017-161-1781

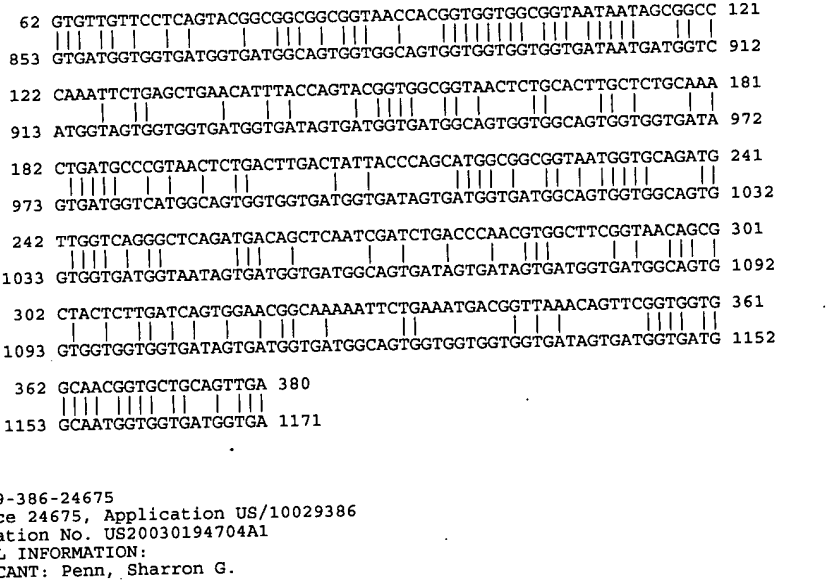
Query Match 8.6%; Score 39; DB 14; Length 1390;
Best Local Similarity 45.1%; Pred. No. 0.021;
Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;





RESULT 3
US-10-292-798-1437
; Sequence 1437, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1437
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1390)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1190)
US-10-292-798-1437

Query Match 8.6%; Score 39; DB 15; Length 1390;
Best Local Similarity 45.1%; Pred. No. 0.021;
Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;



RESULT 4
US-10-029-386-24675
; Sequence 24675, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.

```
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24675
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP001754.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
;   OTHER INFORMATION: EST_HUMAN HIT: BI015471.1, EVALUE 1.00e-03
;   OTHER INFORMATION: NT HIT: AE000415.1, EVALUE 9.30e+00
US-10-029-386-24675

  Query Match             8.4%;  Score 38.2;  DB 14;  Length 625;
  Best Local Similarity   46.2%;  Pred. No. 0.026;
  Matches  162;  Conservative    0;  Mismatches  188;  Indels    1;  Gaps    1;

Qy         29 CAGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCG 88
              || ||  | ||  |  .  |||| |      |  |||| || |  | |  | || ||| |||
Db         87 CATCACCCTTACTTATGAGTAGTGGGGGCGGTGGTGGTGGATGGTGATGATGGTGGCTGCG 146

Qy         89 GTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGT 148
               |    | ||||||||| ||| ||  ||| ||    | |  |||  |     | |   ||
Db        147 ATGGTGATGGTGGTGGTGGTGATGGTAGTGGTGGTAGTGGTGATGTTGGTGATGA-TGGT 205

Qy        149 ACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTA 208
                |||| ||| | | ||    || |         |||||   |||    ||        |
Db        206 GATGTGGTGGTGATTGTGATGGTGGTGGTGTTGGTGATGGTGGTAGTGATGGTGATGATG 265

Qy        209 TTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAA 268
                |    |   ||  ||| |||||||    | || || |   ||    ||||  ||
Db        266 ATGGTGATTCTGATGGCTGTAATGGAAGTGGTGGTGATGGTGGTGGTGATGGTAGTGGTG 325

Qy        269 TCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAA 328
                ||| |          ||    | ||    | | |   |  || | |  | |  ||
Db        326 GTGATGTTGATGGTGATGATGGCAGTGGTGGTGGTGGTGGTGGTGATGGTAGTGGTGGTG 385

Qy        329 ATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTG 379
              ||  |||        |||          ||||||||  | |||| ||  | ||
Db        386 ATGTTGATGGTGATGTTGGCAGTGGTGGTGGTGGTGATGGTGGTGATGGTG 436


RESULT 5
US-10-017-161-1981/c
; Sequence 1981, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1981
;   LENGTH: 1972
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1972)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1772)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (975)..(994)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (1319)..(1328)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1981

  Query Match             8.2%;  Score 37.4;  DB 14;  Length 1972;
  Best Local Similarity   49.2%;  Pred. No. 0.086;
  Matches   98;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         62 GTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCC 121
              ||| |  |  | |  | || |  || |||  |   | ||||||  | |  | || | ||
Db        460 GTGATTATGATGATGATGGTGATGGTGGTGGCGGTGATGGTGGTGATGGTGATGGTGGTG 401

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
                  | ||            | |  | | ||||| |||  || ||    || |    |
Db        400 ATGGTGGTGGTGGTGGTGGTGATGATGATGGTGGTGGTGGCTGTGATTATGATGATGATG 341

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATG 241
              |||||   ||  |  |||     |  |    |  |||  || |||  |||||  ||||
Db        340 GTGATGGTGGTGGCGGTGATGGTGGTGATGGTGATGGTGGTGGTGGTGGTGGTGATGATG 281

Qy        242 TTGGTCAGGGCTCAGATGA 260
              ||||  ||||  |||  |
Db        280 GTGGTGGTGGCTGTGATTA 262


RESULT 6
US-10-292-798-1629/c
; Sequence 1629, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1629
;   LENGTH: 1972
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1972)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1772)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (975)..(994)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
```

; LOCATION: (1319)..(1328)
; OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-1629

Query Match 8.2%; Score 37.4; DB 15; Length 1972;
Best Local Similarity 49.2%; Pred. No. 0.086;
Matches 98; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```
Qy        62 GTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCC 121
             ||| |  |  | |  | || |  || |||  |  | ||||||| | |  | || | ||
Db       460 GTGATTATGATGATGATGGTGATGGTGGTGGCGGTGATGGTGGTGATGGTGATGGTGGTG 401

Qy       122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
               |  ||          | |  |  | ||||| |||  || ||   ||| |      |
Db       400 ATGGTGGTGGTGGTGGTGATGATGATGGTGGTGGTGGCTGTGATTATGATGATGATG 341

Qy       182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATG 241
             |||||   ||  |  |||       |  |    |  ||| || ||| |||||  ||||
Db       340 GTGATGGTGGTGGCGGTGATGGTGGTGATGGTGATGGTGGTGGTGGTGGTGATGATG 281

Qy       242 TTGGTCAGGGCTCAGATGA 260
             ||||  ||||  ||| |
Db       280 GTGGTGGTGGCTGTGATTA 262
```

RESULT 7
US-09-738-626-1857
; Sequence 1857, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1857
; LENGTH: 2580
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1857

Query Match 8.1%; Score 37; DB 9; Length 2580;
Best Local Similarity 47.8%; Pred. No. 0.13;
Matches 140; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

```
Qy        77 ACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGA 136
             | ||| | ||  ||   | | || |  | |||     || | |  ||| | |||   | |
Db      1046 AGGGTGACGCTGGCGCCGATGGCGAAGACGGCGCAGATGGTGAATCAATTACTGTTGTTA 1105

Qy       137 ACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTG---ATGCCCGTA 193
             |||     | | | | || ||| ||     |||  ||       ||     ||   ||
Db      1106 ACACCAGCAACGATGCTGAAGGCAACACTCTTGTTGAGCTCTCCGATGGCACTGTTATTA 1165

Qy       194 ACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCT 253
             | | | |  | | ||      | |  ||||   |||||| |||||    || | |||
Db      1166 CCATTAATAAGGGTGATAAGGGTGACGCTGGCGCCGATGGTGAAGATGGAAGTAATGGCG 1225

Qy       254 CAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATC 313
             |||||| |  || ||    | | | ||    | |||| |    ||   |     ||| |
Db      1226 CAGATGGCGAGTCCATTACGGTCATCGAAACCCGTTTCGATGCGGACGGCAACACTGTTG 1285

Qy       314 AGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAAC 366
             | |  |  ||       |||||| ||  | ||  ||   |     ||| ||
Db      1286 TGGTGTTCTCGAACGGCACTGAAATCACCATCAACAAGGGTGAGAAGGGCGAC 1338
```

RESULT 8
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 8.1%; Score 37; DB 9; Length 3309400;
Best Local Similarity 47.8%; Pred. No. 4.5;
Matches 140; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

```
Qy        77 ACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGA 136
             | ||| | ||  ||   | | ||| |  | |||     || | |  ||| | |||   | |
Db   1784428 AGGGTGACGCTGGCGCCGATGGCGAAGACGGCGCAGATGGTGAATCAATTACTGTTGTTA 1784369

Qy       137 ACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTG---ATGCCCGTA 193
             |||     | | | | || ||| ||     |||  ||       ||     ||   ||
Db   1784368 ACACCAGCAACGATGCTGAAGGCAACACTCTTGTTGAGCTCTCCGATGGCACTGTTATTA 1784309

Qy       194 ACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCT 253
             | | | |  | | ||      | |  ||||   |||||| |||||    || | |||
Db   1784308 CCATTAATAAGGGTGATAAGGGTGACGCTGGCGCCGATGGTGAAGATGGAAGTAATGGCG 1784249

Qy       254 CAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATC 313
             |||||| |  || ||    | | | ||    | |||| |    ||   |     ||| |
Db   1784248 CAGATGGCGAGTCCATTACGGTCATCGAAACCCGTTTCGATGCGGACGGCAACACTGTTG 1784189

Qy       314 AGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAAC 366
             | |  |  ||       |||||| ||  | ||  ||   |     ||| ||
Db   1784188 TGGTGTTCTCGAACGGCACTGAAATCACCATCAACAAGGGTGAGAAGGGCGAC 1784136
```

RESULT 9
US-10-029-386-25133/c
; Sequence 25133, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:

```
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25133
;   LENGTH: 1493
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP001751.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.6
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
US-10-029-386-25133

  Query Match             7.9%;  Score 35.8;  DB 14;  Length 1493;
  Best Local Similarity   44.2%;  Pred. No. 0.26;
  Matches  148;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy         46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
              |||    | |    |  |||| || |  | |     || || || ||| |     ||||||||
Db        987 GGTGATGGTGATGGTGGTGGTGATGGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGT 928

Qy        106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
              |||  | || |  |      |  ||          |     |    |||||  | |     |
Db        927 GGTGGTGATTGTCGTGGTGGTGTTGGTGGTGGTGATGGTGATGGTGGTGGTGATGGTGAT 868

Qy        166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
              |  | || |     |   || ||   || |   || |     |  |       |||| ||
Db        867 GGTCATGATGGTGATGGTGGTGATGGTGATGGTGGCGGTGTTGGTGGTGGTGATGGTGGT 808

Qy        226 GGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGT 285
              ||| ||||||  | || ||||   ||     | ||  |       | |           ||
Db        807 GGTGATGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGATGGTGGTGATGGTGGT 748

Qy        286 GGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTT 345
              ||  | |||    | | |  |  ||||    |  | ||  |   |  |||       | |
Db        747 GGTGTTGGTGGTGGTGATGGTGGTGATGGTGGTGATGGTGATGGTGGTGATGCTGATGGT 688

Qy        346 AAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGA 380
                  |   |||||||   | |||| ||  | |||
Db        687 GGTGATAGTGGTGGTGATGATGGTGGTGATGGTGA 653


RESULT 10
US-10-425-114-31747/c
; Sequence 31747, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 31747
;   LENGTH: 992
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMFLB73212F11_FLI
US-10-425-114-31747

  Query Match             7.8%;  Score 35.6;  DB 12;  Length 992;
  Best Local Similarity   54.6%;  Pred. No. 0.25;
  Matches   71;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy         45 CGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGG 104
              |||    | ||| |  ||  ||| | ||    ||||||||||||      |||||| ||
Db        471 CGGGTCAGGTCTCGGCGGCCTTGGTGGTCTCGGCGGCGGCGGCGGCCTTGGCGGTGGCGG 412

Qy        105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
              |||   | | ||||| ||   || |  |||  |      |      |  |||||| |
Db        411 CGGCCTTGGTGGCGGCTCAGGATCGGGTCTGGGCGGCAGCGGTGGCCTTGGCGGTGGCAG 352

Qy        165 TGCACTTGCT 174
              ||  || | |
Db        351 TGGGCTCGGT 342


RESULT 11
US-10-108-260A-1242
; Sequence 1242, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1242
;   LENGTH: 2122
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-1242

  Query Match             7.8%;  Score 35.6;  DB 15;  Length 2122;
  Best Local Similarity   46.7%;  Pred. No. 0.37;
  Matches  113;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy         90 TAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTA 149
              || ||| | |       | || |||| |  |   ||||||||  | ||||  || |||||
Db        302 TATCCAAGATCTCAAGGCTAGTAATTGGTGAGAAAATTCTGCACAGAACCATTCCCAGTG 361

Qy        150 CGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTAT 209
              | | | ||   | | |     || | | |||    | || |    | ||  || | |
Db        362 GTGAGACTGTTTTTGTCCTTCAGCCCGGGAAAGGCAGACCTGGCTCCCTCCTTTTATTTG 421

Qy        210 TACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAAT 269
              |  | ||  |      |  | |       |||| | ||     ||| |
Db        422 CAAGCCTCAAAACCATCAAGACAGGCAAAGACACTCGAAGGGCGCGGACCTTCGCTGAGA 481

Qy        270 CGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAA 329
              ||       | | ||||     | |   |||| |||  ||| | || | | ||  || |
Db        482 CGCCACCCACGCAAGTGGAGGAAGCACTTGCGCCACTGCTGAGCGGTTGGAAGGGGAACA 541

Qy        330 TT 331
              |
Db        542 CT 543


RESULT 12
US-10-389-647-226/c
; Sequence 226, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
;  APPLICANT: GREENBERG, E. Peter
;  APPLICANT:  SCHUSTER, Martin
;  APPLICANT:  LOSTROH, Candi
;  TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
;  FILE REFERENCE: UIZ-038CP
;  CURRENT APPLICATION NUMBER: US/10/389,647
```

```
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: 09/653730
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: 60/153022
;  PRIOR FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 710
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 226
;   LENGTH: 3078
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-10-389-647-226

  Query Match             7.8%;  Score 35.6;  DB 12;  Length 3078;
  Best Local Similarity   64.6%;  Pred. No. 0.44;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         30 AGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGG 89
              |||    |||||| |||||  ||  |||   ||||  | ||  ||| | | |||||||||
Db       1387 AGCGGGTGTATTCGCCGGTGCCGGACTGCGCGGTGGCGATCGGCAGGAAGCCGGCGGCGG 1328

Qy         90 TAACCACGGTGGTGGCGGTAAT 111
              | | || ||||  || |   ||
Db       1327 TGATCAGGGTGCCGGTGAGCAT 1306


RESULT 13
US-10-369-493-46798
; Sequence 46798, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46798
;   LENGTH: 4389
;   TYPE: DNA
;   ORGANISM: Bacillus subtilis
US-10-369-493-46798

  Query Match             7.8%;  Score 35.6;  DB 15;  Length 4389;
  Best Local Similarity   50.6%;  Pred. No. 0.52;
  Matches   86;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        247 CAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACT 306
              || |  |  ||  |  || | |  |  ||  |  |    |     ||| |||||||| ||
Db       2221 CATGAATTTGACAAAGGCCCGACGGTGCTTTCTGATTTTCTAAGCGGAAACAGCGCAACC 2280

Qy        307 CTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAAC 366
              |||||| |  |||| || |  | | |||| | |||     | | | |   ||| |||
Db       2281 GTTGATCCGGCGAACCGCTATCACTTTGAAGCGCCGGAATTTCCGATTGTCAGTGCAAAC 2340

Qy        367 GGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCA 416
              |  | ||       |  || |       || || |  |||||   | | ||
Db       2341 GTAGATGTGTCAAATGAGCCCAAGCTGAAATCATTTGTCAAAAAGCCGCA 2390


RESULT 14
US-10-029-386-22627/c
; Sequence 22627, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22627
;   LENGTH: 785
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC007540.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.94
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.73
;   OTHER INFORMATION: EST HUMAN HIT: BE767903.1, EVALUE 1.00e-03
;   OTHER INFORMATION: NT HIT: AJ252124.1, EVALUE 1.20e-02
US-10-029-386-22627

  Query Match             7.7%;  Score 35;  DB 14;  Length 785;
  Best Local Similarity   44.8%;  Pred. No. 0.36;
  Matches  134;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy         79 GGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAAC 138
              || |  || | | |  | ||||  || | | || ||  ||     ||  |||      |
Db        444 GGTGATGGTGATGATGATGGTGACGGTGATGATTATGATGGAGATGATGGTGATGGCGAT 385

Qy        139 ATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCT 198
              | | ||   | | |||  || |   ||   || |    |   || ||  || |   |
Db        384 AATGACGGAGATGATGGTAGTGATGATGGTGATGGTGATGATGATGGTGATGGTGATGAT 325

Qy        199 GACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGAT 258
              ||          |   |  ||  || | ||||| || ||||| || |   ||    | |
Db        324 GATGATGGAGATGATGATGGTGATGGTGATAATGATGGAGATGATGATGGTGGTCATGTT 265

Qy        259 GACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGTGG 318
              ||          |||       |  |||  |||  |   | | |   |  ||| | |
Db        264 GATGATGATGGTGATGGTGATGATGATGGTGATGGTGATGATGATGATGGAGATGATGGT 205

Qy        319 AACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTGCTGCAGT 377
              | ||| | |||  || |       | | |  |    |||| ||| | | ||   ||||
Db        204 GATGGCGATAATGATGGAGATGATGGTGATGATGATGGTGATGGTGATGATGTGACAGT 146


RESULT 15
US-10-029-386-7202
; Sequence 7202, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7202
;   LENGTH: 532
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC003693.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
```

```
;   OTHER INFORMATION: NT HIT: AF143426.1, EVALUE 5.10e-01
;   OTHER INFORMATION: SWISSPROT HIT: P31638, EVALUE 7.50e-01
US-10-029-386-7202

  Query Match             7.6%;  Score 34.6;  DB 14;  Length 532;
  Best Local Similarity   48.3%;  Pred. No. 0.41;
  Matches   97;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              |||| || |  | |  | |     |  ||| |  | |||||||| |  ||| ||   ||
Db        294 GGTGGTGGTGGTGATGATGATATTGATGGTCATGATGGTGGTGGTGACAATGATCATGGG 353

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              | ||||        |||   |  | ||||||||||     ||    || |  |||
Db        354 GACCGTGCTGATGGCGGTGTTTGTGATGATGGTGGCGGTGGTGGTGATGGTGGTGGGCAT 413

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              | |    |      |||    | ||       |  | || || | | || ||| | |
Db        414 GAGGGTAGTTGGGGTGGTGATGGCATTACCGTTGATGAGGGTGGTGATGATAGTGATGGT 473

Qy        241 GTTGGTCAGGGCTCAGATGAC 261
              | ||||   ||    ||||||
Db        474 GATGGTGGTGGGAATGATGAC 494


Search completed: March 19, 2004, 03:48:48
Job time : 263.734 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-3	Sequence 3, Appli
2	453	99.3	453	92	US-60-352-946-1	Sequence 1, Appli
3	453	99.3	453	101	US-60-444-371-1	Sequence 1, Appli
4	452.8	99.3	675	13	US-08-978-878-3	Sequence 3, Appli
5	452.8	99.3	675	31	US-09-741-873B-3	Sequence 3, Appli
6	396	86.8	396	13	US-08-978-878-1	Sequence 1, Appli
7	360	78.9	360	31	US-09-741-873B-1	Sequence 1, Appli
8	276.8	60.7	477	17	US-09-252-691-172	Sequence 172, App
9	276.8	60.7	477	17	US-09-252-691C-172	Sequence 172, App
10	276.8	60.7	477	51	US-10-417-886-172	Sequence 172, App
11	254.4	55.8	456	6	US-08-233-642A-56	Sequence 56, Appl
12	254.4	55.8	456	23	US-09-543-407-1	Sequence 1, Appli
13	238.4	52.3	456	23	US-09-543-407-23	Sequence 23, Appl
14	236.8	51.9	456	23	US-09-543-407-19	Sequence 19, Appl
15	230.4	50.5	456	23	US-09-543-407-17	Sequence 17, Appl
16	230.4	50.5	456	23	US-09-543-407-21	Sequence 21, Appl
17	228.8	50.2	456	23	US-09-543-407-27	Sequence 27, Appl
18	227.2	49.8	456	23	US-09-543-407-29	Sequence 29, Appl
19	225.6	49.5	456	23	US-09-543-407-13	Sequence 13, Appl
20	222.4	48.8	456	23	US-09-543-407-15	Sequence 15, Appl
21	220.8	48.4	456	23	US-09-543-407-25	Sequence 25, Appl
22	216	47.4	456	23	US-09-543-407-11	Sequence 11, Appl
23	173.4	38.0	361	6	US-08-233-642A-54	Sequence 54, Appl
24	100	21.9	100	45	US-10-146-492B-77	Sequence 77, Appl
25	100	21.9	100	45	US-10-146-492B-78	Sequence 78, Appl
26	100	21.9	100	45	US-10-146-492B-79	Sequence 79, Appl
27	81.8	17.9	522	17	US-09-252-691-170	Sequence 170, App
28	81.8	17.9	522	17	US-09-252-691C-170	Sequence 170, App
29	81.8	17.9	522	51	US-10-417-886-170	Sequence 170, App
30	43.2	9.5	1083	1	PCT-US01-08631-12549	Sequence 12549, A
c 31	42.6	9.3	37551	81	US-60-245-201-6	Sequence 6, Appli
c 32	42.6	9.3	37551	82	US-60-258-252-2	Sequence 2, Appli
c 33	42.6	9.3	135259	1	PCT-US02-25766-8350	Sequence 8350, Ap
c 34	42.6	9.3	135259	47	US-10-240-425-1585	Sequence 1585, Ap
c 35	42.6	9.3	135259	99	US-60-423-586-24	Sequence 24, Appl
c 36	42.6	9.3	135259	99	US-60-427-194-24	Sequence 24, Appl
37	41.6	9.1	561	17	US-09-252-691-171	Sequence 171, App
38	41.6	9.1	561	17	US-09-252-691C-171	Sequence 171, App
39	41.6	9.1	561	51	US-10-417-886-171	Sequence 171, App
40	39	8.6	1390	43	US-10-017-161-1781	Sequence 1781, Ap
41	39	8.6	1390	48	US-10-292-798-1437	Sequence 1437, Ap
c 42	39	8.6	2730	1	PCT-US01-08631-23586	Sequence 23586, A
43	38.2	8.4	625	43	US-10-029-386-24675	Sequence 24675, A
44	38.2	8.4	32768	77	US-60-213-177-372	Sequence 372, App
45	38.2	8.4	87681	82	US-60-258-275-239	Sequence 239, App

ALIGNMENTS

RESULT 1
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. Coli

US-09-543-407-3

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.9e-130;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

RESULT 2
US-60-352-946-1
; Sequence 1, Application US/60352946
; GENERAL INFORMATION:
; APPLICANT: Hultgren, Scott
; APPLICANT: Matt, Chapman
; TITLE OF INVENTION: BACTERIAL MODEL SYSTEM FOR AMYLOID FORMATION
; FILE REFERENCE: WSHU 2059
; CURRENT APPLICATION NUMBER: US/60/352,946
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Escherichia coli
US-60-352-946-1

Query Match 99.3%; Score 453; DB 92; Length 453;
Best Local Similarity 100.0%; Pred. No. 7.5e-129;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
```

```
Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTAC 453
              |||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTAC 453
```

RESULT 3
US-60-444-371-1
; Sequence 1, Application US/60444371
; GENERAL INFORMATION:
; APPLICANT: Hultgren, Scott J
; APPLICANT: Matt, Chapman R
; TITLE OF INVENTION: BACTERIAL MODEL SYSTEM FOR AMYLOID FORMATION
; FILE REFERENCE: WSHU 2059.1
; CURRENT APPLICATION NUMBER: US/60/444,371
; CURRENT FILING DATE: 2003-01-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Escherichia coli
US-60-444-371-1

Query Match 99.3%; Score 453; DB 101; Length 453;
Best Local Similarity 100.0%; Pred. No. 7.5e-129;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTAC 453
              |||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTAC 453
```

RESULT 4
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-08-978-878-3

Query Match 99.3%; Score 452.8; DB 13; Length 675;
Best Local Similarity 99.6%; Pred. No. 9.9e-129;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||| |||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

```
RESULT 5
US-09-741-873B-3
; Sequence 3, Application US/09741873B
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873B
;  CURRENT FILING DATE:  2003-04-04
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-09-741-873B-3

  Query Match            99.3%;  Score 452.8;  DB 31;  Length 675;
  Best Local Similarity  99.6%;  Pred. No. 9.9e-129;
  Matches  454;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

```
RESULT 6
US-08-978-878-1
; Sequence 1, Application US/08978878
; GENERAL INFORMATION:
;  APPLICANT: NORMARK, Staffan
;  APPLICANT: OLSEN, Arne
;  TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
;  FILE REFERENCE: 012889-081
;  CURRENT APPLICATION NUMBER: US/08/978,878
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: SE 8801723-1
;  EARLIER FILING DATE: 1988-05-06
;  EARLIER APPLICATION NUMBER: US 07/347,189
;  EARLIER FILING DATE: 1989-05-04
;  EARLIER APPLICATION NUMBER: US 07/789,437
;  EARLIER FILING DATE: 1991-11-06
;  EARLIER APPLICATION NUMBER: US 07/970,846
;  EARLIER FILING DATE: 1992-11-03
;  EARLIER APPLICATION NUMBER: US 08/187,865
;  EARLIER FILING DATE: 1994-01-28
;  EARLIER APPLICATION NUMBER: US 08/318,519
;  EARLIER FILING DATE: 1994-10-05
;  EARLIER APPLICATION NUMBER: US 08/495,959
;  EARLIER FILING DATE: 1995-06-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-08-978-878-1

  Query Match            86.8%;  Score 396;  DB 13;  Length 396;
  Best Local Similarity  100.0%;  Pred. No. 3e-111;
  Matches  396;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

```
RESULT 7
US-09-741-873B-1
; Sequence 1, Application US/09741873B
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
```

```
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873B
;  CURRENT FILING DATE:  2003-04-04
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 360
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873B-1

  Query Match             78.9%;  Score 360;  DB 31;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 4e-100;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360


RESULT 8
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match             60.7%;  Score 276.8;  DB 17;  Length 477;
  Best Local Similarity   75.4%;  Pred. No. 2.4e-74;
  Matches  344;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||| ||||| | ||  ||||||||||||||| | || || || ||| | |||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||  || || || |  |||||||||          ||  | || ||    |||| ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||     |||| ||||||||||||||| |||||||||| || |||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              || || |||||  | ||||| ||||| |||||||||||||| || ||||| || ||||||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||| |||||||| |||||||| || |||||||||    |||   ||||| |||||||||
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || || |||||  ||||| | ||||  || |  || ||||||||||| ||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |||||||| || || || |||||||| || ||   |||  | || |  |||   ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||| ||||||||||||||| || |||||||  |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


RESULT 9
US-09-252-691C-172
; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-172

  Query Match             60.7%;  Score 276.8;  DB 17;  Length 477;
  Best Local Similarity   75.4%;  Pred. No. 2.4e-74;
  Matches  344;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||| ||||| | ||  ||||||||||||||| | || || || ||| | |||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||  || || || |  |||||||||          ||  | || ||    |||| ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||     |||| ||||||||||||||| |||||||||| || |||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
```

```
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||| |||||||| |||||||| || |||||||||   |||   ||||| |||||||||
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || || |||||  ||||| | ||||  || |  || ||||||||||| ||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |||||||| || || || |||||||| || ||   |||  | || |  |||   ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||| |||||||||||||| || |||||||
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477
```

RESULT 10
US-10-417-886-172
; Sequence 172, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/10/417,886
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US/09/252,691C
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-10-417-886-172

Query Match 60.7%; Score 276.8; DB 51; Length 477;
Best Local Similarity 75.4%; Pred. No. 2.4e-74;
Matches 344; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||| ||||| | ||  |||||||||||||| | || || || ||| | |||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||  || || || |  |||||||||          ||  | || ||    |||| ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||    |||| |||||||||||||| |||||||||| || |||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              || || |||||  | ||||| ||||| |||||||||||||| || |||||| || ||||||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||| |||||||| |||||||| || |||||||||   |||   ||||| |||||||||
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || || |||||  ||||| | ||||  || |  || ||||||||||| ||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |||||||| || || || |||||||| || ||   |||  | || |  |||   ||||||
```

```
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||| |||||||||||||| || |||||||
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477
```

RESULT 11
US-08-233-642A-56
; Sequence 56, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-642A-56

Query Match 55.8%; Score 254.4; DB 6; Length 456;
Best Local Similarity 72.4%; Pred. No. 2e-67;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||   || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
```

```
Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             55.8%;  Score 254.4;  DB 23;  Length 456;
  Best Local Similarity   72.4%;  Pred. No. 2e-67;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             52.3%;  Score 238.4;  DB 23;  Length 456;
  Best Local Similarity   70.2%;  Pred. No. 1.7e-62;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| | |||  |  |||||  |   | | |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                |   ||    |  | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
```

```
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             51.9%;  Score 236.8;  DB 23;  Length 456;
  Best Local Similarity   70.0%;  Pred. No. 5.4e-62;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  ||||||||||||||| ||     || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||    | ||| |||||| |||||||||   || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||||  | |||  |    ||||||    ||     |  |  |||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              |  || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||   || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 15
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             50.5%;  Score 230.4;  DB 23;  Length 456;
  Best Local Similarity   69.1%;  Pred. No. 5.2e-60;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  ||||||||||||||| ||     || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||    | ||| |||||| |||||||||   || |||| |||| ||||||||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||||  | |||  |    ||||||    ||     |  |  |||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||   || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


Search completed: March 19, 2004, 03:01:41
Job time : 2652.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	452.8	99.3	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	396	86.8	396	5	US-09-741-873C-1	Sequence 1, Appli
c	3	34.2	7.5	79977	6	US-10-765-790-58	Sequence 58, Appl
	4	34.2	7.5	186449	6	US-10-021-698A-715	Sequence 715, App
c	5	33	7.2	439	6	US-10-021-698A-1006	Sequence 1006, Ap
	6	32	7.0	1355	6	US-10-045-674A-593	Sequence 593, App
c	7	30.8	6.8	290892	6	US-10-767-471-10809	Sequence 10809, A
c	8	30.4	6.7	945	6	US-10-767-701-7789	Sequence 7789, Ap
c	9	30.2	6.6	63693	7	US-60-548-091-5635	Sequence 5635, Ap
	10	30.2	6.6	290892	6	US-10-767-471-10809	Sequence 10809, A
c	11	30	6.6	372	6	US-10-021-698A-1217	Sequence 1217, Ap
	12	30	6.6	992	1	PCT-US04-05654-1019	Sequence 1019, Ap
c	13	30	6.6	9278	6	US-10-021-698A-796	Sequence 796, App
	14	29.6	6.5	1203	1	PCT-US04-02242-43	Sequence 43, Appl
	15	29.2	6.4	306	6	US-10-100-683-13342	Sequence 13342, A
	16	29.2	6.4	306	6	US-10-100-683-13343	Sequence 13343, A
	17	29.2	6.4	851	6	US-10-767-795-1598	Sequence 1598, Ap
	18	29.2	6.4	2714	6	US-10-152-372-79	Sequence 79, Appl
	19	29.2	6.4	2776	6	US-10-100-683-3409	Sequence 3409, Ap
c	20	29.2	6.4	4660	6	US-10-788-792-80	Sequence 80, Appl
	21	29	6.4	507	1	PCT-US04-05654-1128	Sequence 1128, Ap
c	22	29	6.4	551	6	US-10-767-701-681	Sequence 681, App
	23	29	6.4	753	6	US-10-767-701-9504	Sequence 9504, Ap
c	24	29	6.4	1350	1	PCT-US04-05654-863	Sequence 863, App
	25	28.8	6.3	6705	5	US-09-032-438C-5	Sequence 5, Appli
	26	28.8	6.3	6705	6	US-10-336-215A-5	Sequence 5, Appli
	27	28.8	6.3	6705	6	US-10-336-219A-5	Sequence 5, Appli
	28	28.8	6.3	6705	6	US-10-340-097B-5	Sequence 5, Appli
	29	28.8	6.3	6819	5	US-09-032-438C-2	Sequence 2, Appli
	30	28.8	6.3	6819	6	US-10-336-215A-2	Sequence 2, Appli
	31	28.8	6.3	6819	6	US-10-336-219A-2	Sequence 2, Appli
	32	28.8	6.3	6819	6	US-10-340-097B-2	Sequence 2, Appli
	33	28.8	6.3	7318	6	US-10-783-528-12	Sequence 12, Appl
	34	28.8	6.3	7488	5	US-09-032-438C-1	Sequence 1, Appli
	35	28.8	6.3	7783	6	US-10-336-215A-1	Sequence 1, Appli
	36	28.8	6.3	7783	6	US-10-336-219A-1	Sequence 1, Appli
	37	28.8	6.3	7783	6	US-10-340-097B-1	Sequence 1, Appli
c	38	28.6	6.3	435	6	US-10-021-698A-1216	Sequence 1216, Ap
	39	28.6	6.3	704	1	PCT-US04-05654-1022	Sequence 1022, Ap
	40	28.6	6.3	860	1	PCT-US04-05654-1018	Sequence 1018, Ap
	41	28.6	6.3	906	1	PCT-US04-05654-1020	Sequence 1020, Ap
	42	28.6	6.3	74279	6	US-10-765-790-16	Sequence 16, Appl
c	43	28.4	6.2	524	6	US-10-767-701-22290	Sequence 22290, A
	44	28.2	6.2	654	6	US-10-767-701-4153	Sequence 4153, Ap
c	45	28.2	6.2	1252	1	PCT-US04-05654-877	Sequence 877, App

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 99.3%; Score 452.8; DB 5; Length 675;
Best Local Similarity 99.6%; Pred. No. 2.4e-145;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||| |||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 86.8%; Score 396; DB 5; Length 396;
Best Local Similarity 100.0%; Pred. No. 5.4e-126;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-10-765-790-58/c
; Sequence 58, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
; LENGTH: 79977
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-58

Query Match 7.5%; Score 34.2; DB 6; Length 79977;
Best Local Similarity 48.2%; Pred. No. 0.61;
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```
Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              |||| || |  | |  | |  |  || | | |  | || | ||| | |  | || | ||
Db      15147 GGTGGTGATGGTGATGATGATGATGGTGATGATGAAGGAGATGGTGATGGTGATGGTGGT 15088

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
                 ||  |||     |   |    || | |||| |||  |   || |   | |     |
Db      15087 GATGATGATGATGGTGATGATGGTGAGGATGGTGATGGTGATGATGGAGGAGATGGTGAC 15028

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
                |||   || |   |||     |  |     |   ||  || ||| | | ||  |||
Db      15027 GGCGATGGTGGTGATGGTGATCATGATGGTGATGATGGTGATGGTGGTGAGGATGGTGAT 14968

Qy        241 GTTGGTCAGGGCTCAGATG 259
              | |||| | ||   |||||
Db      14967 GGTGGTGATGGTGAAGATG 14949
```

RESULT 4
US-10-021-698A-715
; Sequence 715, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL

```
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 715
;    LENGTH: 186449
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (40991)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (41047)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (41057)..(41156)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (70192)..(70291)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90076)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90104)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90234)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90262)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90290)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90306)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90313)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90316)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (90344)..(90443)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (93719)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (111786)..(111885)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (129353)..(129452)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (144859)..(144958)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (155513)..(155612)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (166141)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (166388)..(166487)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (172617)..(172716)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (177984)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (178024)..(178123)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (180005)..(180104)
;    OTHER INFORMATION: a, t, c or g
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (183290)..(183389)
;    OTHER INFORMATION: a, t, c or g
US-10-021-698A-715

  Query Match             7.5%;  Score 34.2;  DB 6;  Length 186449;
  Best Local Similarity   43.9%;  Pred. No. 0.91;
  Matches  147;  Conservative    0;  Mismatches  188;  Indels    0;  Gaps    0;

Qy         46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
              |||   | |   |  | || || |  | |  | || |  |  ||| |   | |||||
Db     148605 GGTGATGATGGTGGTGATGGTGATGATGATGATGGTGATGATGGTGATGCTGATGGTGAT 148664

Qy        106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
              | | || || | |       ||  ||       | | | ||    | ||||  | |     |
Db     148665 GATGATGATGGTGATGGTGATGATGGTGATGATAGTGACTGTGATGGTGATGATGGTGAT 148724

Qy        166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
              |    || |         || ||   || ||| |||       | |    |   || ||
Db     148725 GGTGGTGATGATGGTGATGGTGATGGTGACTGTGATGGTGATGGTGGTGATGGTGATGGT 148784

Qy        226 GGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGT 285
              ||| ||| ||  |||| |||| | |     | |||          |||      |  ||
Db     148785 GGTGATGATGGTGATGATGGTGATGCTGATGGTGATGATGATGGTGATGGTGATGATGGT 148844

Qy        286 GGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTT 345
              |     ||| || | | |  | ||||          ||      ||  |||       | |
Db     148845 GATGATGGTGACTGTGATGGTGATGATGGTGATGGTGGTGGTGATGATGATGGTGATGAT 148904

Qy        346 AAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGA 380
              |    | ||||||  | |||| ||  | |||
Db     148905 GGTGATGGTGATGGTGGTGATGGTGATGATGGTGA 148939


RESULT 5
US-10-021-698A-1006/c
```

; Sequence 1006, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4044US1
; CURRENT APPLICATION NUMBER: US/10/021,698A
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/211,749
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 6160
; SOFTWARE: Patentin 2.1
; SEQ ID NO 1006
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-021-698A-1006

Query Match 7.2%; Score 33; DB 6; Length 439;
Best Local Similarity 47.8%; Pred. No. 0.14;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy          46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
               ||| | | |    | ||||  || |  |     | || |  || ||| |    || || |||
Db         314 GGTGGTGATGGTGGTGGTGGTGATGGTGGTGATGGTGATGGTGGTGATGGCGATGATGGT 255

Qy         106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
               | | ||  |   |||    ||  |||     | | |    | ||| | ||    |
Db         254 GATGATGGTGATGGCGATGATGGTGATAGTGATGGTGATGATGGTGATGGTGATAGTGAT 195

Qy         166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
               |   ||  |   |  || ||  || |  |||    | |  |    |  ||||  |
Db         194 GATAATGGTGATGATGATGGTGATGGTGATAGTGATGATAATGGTGATGATGATGGTGAT 135

Qy         226 GGTAATGGTGCAGATGTTGGT 246
               |||| |||||  | |  ||||
Db         134 GGTAGTGGTGATGGTAGTGGT 114

RESULT 6
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
US-10-045-674A-593

Query Match 7.0%; Score 32; DB 6; Length 1355;
Best Local Similarity 25.4%; Pred. No. 0.51;
Matches 59; Conservative 64; Mismatches 109; Indels 0; Gaps 0;

```
Qy   87 CGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCA 146
               :||:|| : |:| : |:| : . : |::::||: ||: || | : |:| : |
Db   663 YGGYAARTTYMGWGAYTGYGCYTTYCAYWSYGGYTTYAAYGARGAYCCWTTYGTYTGYGA 722

Qy   147 GTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGAC 206
               :||: ||:   : :::| :| |:| :| : |:| : : : : :| ::
Db   723 RTAYCARGGYCARWSKWSYGAYYTRCCKCARCCWCCKGTYAAYGCKGGYGGYGGYWSYGG 782

Qy   207 TATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTC 266
               : : :   :||:||: :: | ||:| :| : :| ||:   :| |:
Db   783 YGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGGYGGYWSYGGRGGYGG 842

Qy   267 AATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGTGG 318
               : :| :   : | :| :| :| || | ||:| : : |: |: |:
Db   843 YWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAAYAARGGS 894

RESULT 7
US-10-767-471-10809/c
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
; LENGTH: 290892
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(290892)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

Query Match 6.8%; Score 30.8; DB 6; Length 290892;
Best Local Similarity 46.3%; Pred. No. 16;
Matches 101; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

```
Qy          42 CTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGG 101
               || |  ||| | | |    || || |  | |  | |  || | |  ||  |   ||
Db     234185 CTACATTAGTGATGTTGGTAATGGTGGTGGTGATGATGATGGTGATGCCAATGATTATGA 234126

Qy          102 TGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAA 161
                ||  | | ||  ||  ||     ||  |||      | | ||||       | ||  |||||
Db     234125 TGATGATGATGGTAATGGTGATGATGGTGATGGTGATAGTTACAGTAGTGATGTTGGTAA 234066

Qy          162 CTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGG 221
                |   || |    |  ||||||   | | | |||       |  |       ||||
Db     234065 TGATAGTGGTGGTGACGACGGTGATGCTAATGATTATGATGATGATGGTGGTGGTAATGG 234006

Qy          222 CGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATG 259
                | | | |  ||  | || || ||  |||| ||||
Db     234005 TGATGATGGTAATGTTGGTGATGATCGTGGCTGTGATG 233968

RESULT 8
```

```
US-10-767-701-7789/c
; Sequence 7789, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 7789
;   LENGTH: 945
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS14531_1
US-10-767-701-7789

  Query Match             6.7%;  Score 30.4;  DB 6;  Length 945;
  Best Local Similarity   53.3%;  Pred. No. 1.5;
  Matches   64;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        133 CTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGT 192
              ||| | |    |  | | || |  |||  |  || ||||| ||| ||| ||  ||    |
Db        569 CTGGAAGTAATCTTGCAAGGCGAGGGTCTCAATGGACTTGTTCTTCAAGCTCTTGATGGC 510

Qy        193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGC 252
              | || ||     ||    |||   || ||  | | |||| |       ||| ||  |||||
Db        509 ATCTATGGAGGCACGTGCACCATCCACGGTTGTGATAATTGGAACCTTGTAGGCGAGGGC 450


RESULT 9
US-60-548-091-5635/c
; Sequence 5635, Application US/60548091
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STROKE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001506
;  CURRENT APPLICATION NUMBER: US/60/548,091
;  CURRENT FILING DATE:  2004-02-27
;  NUMBER OF SEQ ID NOS: 24433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5635
;   LENGTH: 63693
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-548-091-5635

  Query Match             6.6%;  Score 30.2;  DB 7;  Length 63693;
  Best Local Similarity   46.8%;  Pred. No. 13;
  Matches   95;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              | || |  |  | |  | || || |  |||       |||||| ||| || || | ||
Db      53590 GATGGTTATGGTGATGATGGTGGTGATGGTGGTGGTAGTGGTGGTGGTGATGATGGTGGT 53531

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
                ||  ||       |   | |     | |||||| || |     |   |     |  |
Db      53530 GGTGATAATGGTAGTGATGATAATGGTGATGGTGGTGGAATGGTGGTGATGATGGTTGAT 53471

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              | |||||     |     |||       |  |     |   ||  ||| |||||| ||  | |
Db      53470 AATGATGGTGATGGTGGTGATGATGGTGATGGTGATGTTGATGGCAGTAATGATGGGGGT 53411

Qy        241 GTTGGTCAGGGCTCAGATGACAG 263
              | ||||   ||    ||||| ||
Db      53410 GGTGGTGGTGGTGGTGATGATAG 53388


RESULT 10
US-10-767-471-10809
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
;   LENGTH: 290892
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(290892)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-767-471-10809

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 290892;
  Best Local Similarity   49.7%;  Pred. No. 26;
  Matches   77;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          97 GGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGC 156
               |||||||   || | || | ||       | | |         | ||       | |||
Db      257220 GGTGGTGGTAATAGTGATGGTGGTGATGGTGGTAATGATGGGGGTGATGGTGTTGATGGT 257279

Qy         157 GGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAG 216
               ||| ||  |||  ||| |  |   | |||||   || |   | ||   | |  |||  |
Db      257280 GGTGACGGTGCTGTTGGTGGGGGGAATGATGGTGGTGATAGTAACGGAAATCATACTGAT 257339

Qy         217 CATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGG 251
               || || ||| || ||    ||||||  | | ||
Db      257340 GCAGGTGGTGGTGATTGTAGTCATGTTGATGATGG 257374


RESULT 11
US-10-021-698A-1217/c
; Sequence 1217, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 1217
;   LENGTH: 372
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-021-698A-1217

  Query Match             6.6%;  Score 30;  DB 6;  Length 372;
  Best Local Similarity   43.1%;  Pred. No. 1.3;
  Matches  144;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy          62 GTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCC 121
               ||| || |  | |  | || || |    || |     | || ||| ||| ||| |   ||
Db         340 GTGATGGTGATGATGATGGTGGTGATAGTGATGGTGATGATGGTGGTGATAGTGATGGTG 281
```

```
Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
Db        280 ATGATGGTGATAGTGATGATGGTGGTTATAGTGATGGTGATGATGGTGGTGATAGTGATG 221

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATG 241
Db        220 GTGATGATAGTGATGATAGTGATGGTGATGATGATGGTGGTGATAGTGATGGTGATGATG 161

Qy        242 TTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCG 301
Db        160 ATGGTGGTAATACTGATGATGGTGCTGGTGATGATGGTGGTGATAGTGATGGTGATGATG 101

Qy        302 CTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTG 361
Db        100 GTGATAGTGATGATGGTGGCGATAGTGATGGTGATGATGGTGGTGATAGTGGTGGTGATG 41

Qy        362 GCAACGGTGCTGCAGTTGACCAGACTGCATCTAA 395
Db         40 ATGATGGTGGTAATAGTGATGATGGTGATAGTAA 7
```

RESULT 12
PCT-US04-05654-1019
; Sequence 1019, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Gutterson, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E
; APPLICANT: Kumimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1019
; LENGTH: 992
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (978)..(984)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (987)..(987)
; OTHER INFORMATION: n is a, c, g, or t
PCT-US04-05654-1019

Query Match 6.6%; Score 30; DB 1; Length 992;
Best Local Similarity 61.5%; Pred. No. 2.1;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy         42 CTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGG 101
Db        324 CTATGGCGGGGGACGAGGAGGTGGTGGAGGTGGATATGGCGGTGGCGGCGGCTACGGTGG 383

Qy        102 TGGCGGTAATAATAGCGG 119
Db        384 TGGCGGTGGCTATGGTGG 401
```

RESULT 13
US-10-021-698A-796/c
; Sequence 796, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4044US1
; CURRENT APPLICATION NUMBER: US/10/021,698A
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/211,749
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 6160
; SOFTWARE: Patentin 2.1
; SEQ ID NO 796
; LENGTH: 9278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-021-698A-796

Query Match 6.6%; Score 30; DB 6; Length 9278;
Best Local Similarity 46.3%; Pred. No. 6.1;
Matches 99; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```
Qy         46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
Db        359 GGTGGTGATGATGGTGATGATGATGGTGACGGTGGTGATGGTGATGGTGATGATGGTGAT 300

Qy        106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
Db        299 GATGGTGACGGTGGTGATGGTGCTGATGATGATGGTGATGCTGATGGTGATGGTGACGGT 240

Qy        166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
Db        239 GATGATGATGGTGACGGTGATGATGGTGATGATGGTGATGGTGATGCTGATGGTGGTGGT 180

Qy        226 GGTAATGGTGCAGATGTTGGTCAGGGCTCAGATG 259
Db        179 GGTGATGATGGTGGTGATGATGATGATGATGATG 146
```

RESULT 14
PCT-US04-02242-43
; Sequence 43, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; CURRENT FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 43
;   LENGTH: 1203
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-43

  Query Match             6.5%;  Score 29.6;  DB 1;  Length 1203;
  Best Local Similarity   56.0%;  Pred. No. 3.2;
  Matches   56;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy         36 CGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCA 95
              |||   | |||   ||   |       | ||||  ||| || | | || ||||   | | |
Db        228 CGTGAACACCGTGCGCCTGCGCCTGTGGGTTGACCCTAAGGATGCCGACGGCAACACCTA 287

Qy         96 CGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTG 135
              ||| || ||| ||||||||  ||     |    || ||||
Db        288 CGGCGGCGGCAGTAATAATCTCGAAAACACGATTGCGCTG 327


RESULT 15
US-10-100-683-13342
; Sequence 13342, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13342
;   LENGTH: 306
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-13342

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 306;
  Best Local Similarity   52.5%;  Pred. No. 2.3;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        128 CTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATG 187
              || || |||      |  ||  ||  |||      | | |||| ||| |    || ||  ||
Db        124 CTCAGGTGAGTTCCTCTCACCACCCTGGGCAGCAGTGTGCAGTTGGTTACAAAGCTCTTG 183

Qy        188 CCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTC 247
               |   |   ||||  ||||   |   | | ||| |||    |  || || || |||   ||
Db        184 TCTTTTGTTCTGGATTGAAAGTCCGCAAACATAGCGTTTATGCTGTTGAAGCTGTGAATC 243

Qy        248 AG 249
              ||
Db        244 AG 245


Search completed: March 19, 2004, 03:10:00
Job time : 57.1445 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No: is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	453.4	99.4	680	12	BJ618688	BJ618688 BJ618688
c	2	60.8	13.3	558	14	CD374421	CD374421 TNWbmfC2H
c	3	43.4	9.5	607	12	BJ348812	BJ348812 BJ348812
	4	38	8.3	282	29	AL758610	AL758610 Arabidops
c	5	37.2	8.2	540	12	BM382763	BM382763 MEST554-D
c	6	37.2	8.2	771	28	BZ997124	BZ997124 PUGGV50TB
	7	37.2	8.2	797	28	BZ997127	BZ997127 PUGGV50TD
	8	37.2	8.2	818	29	CG200536	CG200536 PUIJS94TD
c	9	37	8.1	789	28	BZ115869	BZ115869 CH230-255
	10	36	7.9	1171	13	BX342087	BX342087 BX342087
c	11	35.8	7.9	675	14	CB642466	CB642466 OSJNEb02K
c	12	35.8	7.9	774	29	CNS02APC	AL188841 Tetraodon
	13	35.8	7.9	935	10	BE572585	BE572585 601329541
	14	35.6	7.8	435	10	BE604165	BE604165 WHE1413-1
	15	35.6	7.8	498	10	BE604467	BE604467 WHE1413-1
c	16	35.6	7.8	523	12	BM380389	BM380389 MEST519-C
	17	35.6	7.8	537	28	B64870	B64870 CIT-HSP-201
	18	35.6	7.8	642	28	BZ421645	BZ421645 hz32f09.b
c	19	35.6	7.8	652	12	BM379271	BM379271 MEST503-B
c	20	35.6	7.8	684	29	CC616249	CC616249 OGWEH75TV
	21	35.6	7.8	769	29	CG101505	CG101505 PUIEO88TD
c	22	35.6	7.8	793	29	CG364848	CG364848 OG3AZ85TV
c	23	35.6	7.8	839	28	BZ969045	BZ969045 PUGFX86TD
c	24	35.6	7.8	854	29	CG142634	CG142634 PUFQE81TD
	25	35.6	7.8	867	29	CG142632	CG142632 PUFQE81TB
	26	35.6	7.8	880	28	BZ571017	BZ571017 msh2_168.
	27	35.6	7.8	921	29	CG364838	CG364838 OG3AZ85TH
c	28	35.6	7.8	926	29	CC627745	CC627745 OGVBG10TV
	29	35.6	7.8	926	29	CG211571	CG211571 OGXDF50TV
	30	35.6	7.8	950	28	CC327866	CC327866 OGUEO87TH
	31	35.6	7.8	970	29	CC691748	CC691748 OGOBQ41TV
c	32	35.6	7.8	1245	28	BZ578681	BZ578681 msh2_5964
c	33	35.4	7.8	244	29	CE219585	CE219585 tigr-gss-
	34	35.2	7.7	379	12	BJ266281	BJ266281 BJ266281
	35	35.2	7.7	634	12	BJ266096	BJ266096 BJ266096
	36	35.2	7.7	708	12	BJ272995	BJ272995 BJ272995
	37	35.2	7.7	1251	14	CD509711	CD509711 CDA97-G08
c	38	35	7.7	629	12	BJ404131	BJ404131 BJ404131
	39	34.8	7.6	685	12	BJ249080	BJ249080 BJ249080
c	40	34.6	7.6	582	12	BM276856	BM276856 952012H03
	41	34.6	7.6	586	10	AW759708	AW759708 sl47a01.y
c	42	34.6	7.6	701	14	CD436924	CD436924 EL01N0365
	43	34.6	7.6	818	14	CD098848	CD098848 AGENCOURT
	44	34.6	7.6	978	29	CNS00K6G	AL077284 Drosophil
c	45	34.2	7.5	306	14	R55792	R55792 yg89h05.s1

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.

FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 99.4%; Score 453.4; DB 12; Length 680;
Best Local Similarity 99.6%; Pred. No. 1.5e-114;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
CD374421/c
LOCUS CD374421 558 bp mRNA linear EST 19-AUG-2003
DEFINITION TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf) Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION CD374421
VERSION CD374421.1 GI:31229939
KEYWORDS EST.
SOURCE Wuchereria bancrofti
ORGANISM Wuchereria bancrofti
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Wuchereria.
REFERENCE 1 (bases 1 to 558)
AUTHORS Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
TITLE Toward the characterization of the Wuchereria bancrofti microfilarial transcriptome with comparisons to those of Brugia malayi and Onchocerca volvulus
JOURNAL Unpublished (2003)
COMMENT Contact: Thomas B. Nutman
Laboratory of Parasitic Diseases
NIH
Building 4 Room 126, Bethesda, MD 20892-0425, USA
Email: tnutman@niaid.nih.gov

Seq primer: pBluescript T3.

FEATURES Location/Qualifiers
source 1. .558
/organism="Wuchereria bancrofti"
/mol_type="mRNA"
/db_xref="taxon:6293"
/clone="TNWbmfC2H3"
/dev_stage="microfilaria"
/lab_host="XL1-Blue MRF'"
/clone_lib="Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Lymphatic filarial nematode parasite of humans. mRNA was prepared from approximately 85,000 microfilariae isolated from the blood of an infected individual from Guyana, South America and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 2.9 x 10E5 independent recombinants and the average insert size is ~1kb. The library was constructed by Sandra J. Laney. The library is available from Dr.S.A.Williams, email: genome@smith.edu."

ORIGIN

Query Match 13.3%; Score 60.8; DB 14; Length 558;
Best Local Similarity 77.5%; Pred. No. 6.8e-06;
Matches 93; Conservative 0; Mismatches 25; Indels 2; Gaps 2;

```
Qy         63 TGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAAT-AGCGGCC 121
              |||||   |   |   |||| |||||||||| |||||||||||||||||||| |||||||
Db        174 TGTTGGCACCTGCTGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCC 115

Qy        122 CAAATTCTGAGCTGAACATTTACC-AGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||  |||||   ||  || ||  ||||||||| | |||||| |||||||||| ||
Db        114 CAAATTCTANGCTGACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55
```

RESULT 3
BJ348812/c
LOCUS BJ348812 607 bp mRNA linear EST 06-MAR-2002
DEFINITION BJ348812 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda34o11 3', mRNA sequence.
ACCESSION BJ348812
VERSION BJ348812.1 GI:19219319
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 607)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES Location/Qualifiers
source 1. .607
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda34o11"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"

ORIGIN

Query Match 9.5%; Score 43.4; DB 12; Length 607;

```
  Best Local Similarity   47.7%;  Pred. No. 0.47;
  Matches   95;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              | || || || | |  | | ||| | | | | || || ||| | | || ||   | |
Db        351 GATGGTGATCATNACCATGACGGTGATGATNATGACCATGATGGTGATGATGATNATGAC 292

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              |    | |||     |   | |||| ||||||  | | |   |   | || |    |
Db        291 CATGGTGATGATGATNATNATGACCATGACGGTGATGATGATGATNACCATGATGNTGAT 232

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |||||  || | ||  |||     |     |   ||  |   | ||  ||  |||
Db        231 GNTGATGACCATGACGNTGATGATNATNACCATGANGNTGATGATNATNATCATGGTGAT 172

Qy        241 GTTGGTCAGGGCTCAGATG 259
              | || | |   |   ||||
Db        171 GATGATNATNACCANGATG 153


RESULT 4
AL758610
LOCUS       AL758610                282 bp    DNA     linear   GSS 18-JUN-2002
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-160E03-013216,
            genomic survey sequence.
ACCESSION   AL758610
VERSION     AL758610.1  GI:21496958
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 282)
  AUTHORS   Rosso,M., Strizhov,N., Li,Y. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion close to or within gene At3g28590. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .282
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-160E03-013216"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
ORIGIN

  Query Match             8.3%;  Score 38;  DB 29;  Length 282;
  Best Local Similarity   46.8%;  Pred. No. 10;
  Matches  109;  Conservative    0;  Mismatches  123;  Indels    1;  Gaps    1;

Qy         13 AAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCT 72
              |||  || |   | |||||   || |  | | |  ||      |
Db          1 AAATAAGGAAATAATGCAGATTTCCTTATTTTCCTTATTTTACANNNNNNNNNNNNNNNN 60

Qy         73 CAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAG 132
                  || |  | ||||   || |||| ||||  |||||| |   |||     |||
Db         61 NNNNNNNATGGTGATGATAACTGTGGCGGTGCCGGTGGTAATAGTGAAACAACAGTGGAG 120

Qy        133 CTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGT 192
              ||   ||      | | |||| || ||||  |||   |||          || || | |
Db        121 GTGGCGATGGTGGCG-ATGGTGACGATAACAGTGCCGGTGCCGGTGGTGATGGTGACGGC 179

Qy        193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGG 245
              |||  || |   ||   |  | |   || || |  |||||||  |  | |||
Db        180 AACAATGGCGGTACCGGTGACGATGGTGACGACAATAATGGTGGTGCCGGTGG 232


RESULT 5
BM382763/c
LOCUS       BM382763                540 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  MEST554-D07.univ ISUM6 Zea mays cDNA clone MEST554-D07 3', mRNA
            sequence.
ACCESSION   BM382763
VERSION     BM382763.1  GI:18181553
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Wen,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.
  TITLE     Expressed Sequence Tags from B73 Maize: various stages and tissues
            including seedlings treated with a variety of hormones
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Patrick S. Schnable
            Schnable Laboratory
            Iowa State University
            G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
            Tel: 515-294-0975
            Fax: 515-294-2299
            Email: schnable@iastate.edu
            Individual basecall and confidence value were assigned using the
            Phred software(http://www.phrap.org/). Overall sequence quality
            assessment and vector trimming was conducted using the Lucy
            software (version 1.16s, http://www.tigr.org/softlab/). Lucy
            parameters were set to ensure an overall trimmed quality of 97.5%
            or better without any vector fragments in the chosen high quality
            region of each sequence. Low-quality bases between the poly-T and
            the high-quality region were replaced with N's to serve as spacers
            using a Perl program (est_process.pl), written by Dr. Hui-Hsien
            Chou.
            PCR PRimers
            FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
            BACKWARD: primer Sp6 (ATT TAG GTG ACA CTA TAG)
            Seq primer: universal (GTA AAA CGA CGG CCA GT)
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="MEST554-D07"
                     /tissue_type="mixed"
                     /lab_host="DH10B"
```

/clone_lib="ISUM6"
/note="Vector: pSlip7 (4.43 kb); Site_1: EcoRI; Site_2: NotI; Tissue samples were collected and partially pooled prior to RNA extraction. First-strand cDNAs were prepared from 21 individual pools of oligo-dT selected mRNAs by priming with 21 different NotI oligo-dT tag primers (5'-AACTGGAAGAATTCGCGGCCGCNNNNNNTTTTTTTTTTTTTTTTT-3'). Distinguishable 'bar code' tags, (N)6, were used for each separate first-strand cDNA synthesis. Hence, these bar code tags can be used to identify the mRNA pool from which a particular cDNA clone was derived. The 'bar code' tags associated with specific tissue sources are: ATACGC--Germinated seeds and seedlings (1, 2, 8, 11 DAG); ACTGGC--Mixed mature tissues (17, 21, 38, 69, 77 DAG); CACAGC--Kernels (3, 5, 10, 15, 20, 25, 30, DAP); TAACCC--Adventious roots (65 DAG); CAGGCG--Tassels (3-39 cm, 53 and 56 DAG); AGGTAC--Immature ears (0.2-3.0 cm, 53, 56, 59 DAG); TGAGCG--Husks (73 DAG); GACCAC--Silks; AATCGG--unpollinated first ears; CTAAGG--ear shanks; GAAGAG--etiolated seedlings; AGTGAG--callus; GTGGAC--Cycloheximide-treated callus; GTCACC--Anaerobic treated seedlings; CGTCCA--NAA (a-Naphthalene acetic acid)-treated seedlings; GATGCC--Kinetin-treated seedlings; AAGACC--ACPC (1-aminocyclopropane-1-carboxylix acid)-treated seedlings; GCCTCA--Brassinolide-treated seedlings; CTAGCC--ABA (Abscisic acid)-treated seedlings; TACGGA--GA (Gibberellic acid)-treated seedlings; GCAGGA--JA (Jasmonic acid)-treated seedlings; Equal amounts of first-strand cDNA from each reaction were combined and used as template for DNA PolI-catalyzed 2nd strand synthesis. After the addition of EcoRI adaptors, ds-cDNAs were digested with NotI. Molecules between 0.5 and 2.0 kb were directionally cloned into the EcoRI and NotI sites of the pSlip7 expression vector. Plasmid DNA isolated from the library was digested with NotI to remove empty vector clones. Linear DNAs from 5.4 to 7 kb were gel purified and ligated at low concentration to promote recircularization. Ligation products were precipitated and transformed into DH10B host cells. The complexity of the resulting library was 1.2 x 10X6."

ORIGIN

Query Match 8.2%; Score 37.2; DB 12; Length 540;
Best Local Similarity 55.4%; Pred. No. 23;
Matches 72; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        45 CGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGG 104
             |||   | ||| |  ||  ||| |  ||   ||||||||||||       ||||||  ||
Db       410 CGGGTCAGGTCTCGGCGGCCTTGGTGGTCTCGGCGGCGGCGGCGGCCTTGGCGGTGGCGG 351

Qy       105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
             |||   | | |||||  ||   || |  |||  |   ||    |  |||||||  |
Db       350 CGGCCTTGGTGGCGGCTCAGGATCGGGTCTGGGCGGCAACGGTGGCCTTGGCGGTGGCAG 291

Qy       165 TGCACTTGCT 174
             ||  || | |
Db       290 TGGGCTCGGT 281
```

RESULT 6
BZ997124/c
LOCUS BZ997124 771 bp DNA linear GSS 25-MAR-2003
DEFINITION PUGGV50TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa381I04, genomic survey sequence.
ACCESSION BZ997124
VERSION BZ997124.1 GI:29240541
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 771)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUGGV50TD
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .771
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa381I04"
/clone_lib="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high CoT selected genomic DNA library"
ORIGIN

Query Match 8.2%; Score 37.2; DB 28; Length 771;
Best Local Similarity 50.6%; Pred. No. 27;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy       193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGC 252
             |||||  |     || ||| | | || || || ||||  | ||    ||   |
Db       374 AACACTTCGTCAGTAATAACCGAACGTGTCGTCGATAATAATTCAATGCTTCACCGTTAA 315

Qy       253 TCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGAT 312
             | |    ||   || | |   | ||| | ||| | |||||||| | | | || ||| ||
Db       314 TAATCCAACGTTCCACTAGTAATAACCAAGCGTTTCATCGGTAATAACCCAACGCTTCAT 255

Qy       313 CAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTG 370
             | ||  |   ||  |||  | | |     |||  |||| |||  ||| | |
Db       254 CGGTAATAATCGAACGTTTCGTGAGTAATAACCCAACGCTTCGCCGGTAATAACCGAG 197
```

RESULT 7
BZ997127
LOCUS BZ997127 797 bp DNA linear GSS 25-MAR-2003
DEFINITION PUGGV50TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa381I04, genomic survey sequence.
ACCESSION BZ997127
VERSION BZ997127.1 GI:29240544
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 797)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUGGV50TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .797

```
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa381I04"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             8.2%;  Score 37.2;  DB 28;  Length 797;
  Best Local Similarity   50.6%;  Pred. No. 28;
  Matches   90;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGC 252
              ||| ||   |     || ||| | | || || || ||||  | ||    ||   |
Db        505 AACACTTCGTCAGTAATAACCGAACGTGTCGTCGATAATAATTCAATGCTTCACCGTTAA 564

Qy        253 TCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGAT 312
              | |    ||   || | |   | ||| | |||  | |||||||| | | | || ||| ||
Db        565 TAATCCAACGTTCCACTAGTAATAACCAAGCGTTTCATCGGTAATAACCCAACGCTTCAT 624

Qy        313 CAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTG 370
              | |||   |    ||   |||   | | |       |||  ||||  |||   ||| | |
Db        625 CGGTAATAATCGAACGTTTCGTGAGTAATAACCCAACGCTTCGCCGGTAATAACCGAG 682


RESULT 8
CG200536
LOCUS       CG200536                818 bp    DNA     linear   GSS 21-AUG-2003
DEFINITION  PUIJS94TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0599O19,
            genomic survey sequence.
ACCESSION   CG200536
VERSION     CG200536.1  GI:34091597
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa0599O19"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             8.2%;  Score 37.2;  DB 29;  Length 818;
  Best Local Similarity   50.6%;  Pred. No. 28;
  Matches   90;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGC 252
              ||| ||   |     || ||| | | || || || ||||  | ||    ||   |
Db        277 AACACTTCGTCAGTAATAACCGAACGTGTCGTCGATAATAATTCAATGCTTCACCGTTAA 336

Qy        253 TCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGAT 312
              | |    ||   || | |   | ||| | |||  | |||||||| | | | || ||| ||
Db        337 TAATCCAACGTTCCACTAGTAATAACCAAGCGTTTCATCGGTAATAACCCAACGCTTCAT 396

Qy        313 CAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTG 370
              | ||   |    ||   |||   | | |       |||  ||||  |||   ||| | |
Db        397 CGGTAATAATCGAACGTTTCGTGAGTAATAACCCAACGCTTCGCCGGTAATAACCGAG 454


RESULT 9
BZ115869/c
LOCUS       BZ115869                789 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-255D7.TVB CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-255D7, genomic survey sequence.
ACCESSION   BZ115869
VERSION     BZ115869.1  GI:23756816
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-255D7.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 255  row: D  column: 7
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-255D7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             8.1%;  Score 37;  DB 28;  Length 789;
  Best Local Similarity   45.8%;  Pred. No. 31;
  Matches  127;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy         46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
              ||| | | |   | | || || | |   | || || || ||| |  | | || ||
Db        680 GGTGGTGGTGGTGATGCTGATGATGTTGCTGATGGTGGTGGTGGTGATGATGCTGATGAT 621

Qy        106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
              | |  | || | |       || ||       |   |  |    |||| | |     |
Db        620 GTTGCTGATGGTGCTGGTGATGATGCTGATGATGAGGATGATGGTGGTGATGATGGTGAT 561
```

```
Qy        166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
              |     ||||          ||||||   ||       |||      ||  | | |    |||  ||
Db        560 GGTGATGCTGCTGCTGCTGATGGTGGTGGTGGTGATGCTGCTGCTGCTGATGGTGGTGGT 501

Qy        226 GGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGT 285
              ||| ||| ||| | || || | |   ||     | |||      |     ||| |  |  ||
Db        500 GGTGATGCTGCTGCTGCTGATGGTGGTGGTGGTGATGATGCTGATGATGTTGCTGATGGT 441

Qy        286 GGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACG 322
              ||     ||| |     |||| |   |||| |   | | |
Db        440 GGTGGTGGTGATGATGCTAATGCTGATGATGAGGATG 404
```

RESULT 10
BX342087
LOCUS BX342087 1171 bp mRNA linear EST 01-MAY-2003
DEFINITION BX342087 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens cDNA clone CS0DK009YH08 5-PRIME, mRNA sequence.
ACCESSION BX342087
VERSION BX342087.1 GI:30308886
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1171)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2235.f For more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DK009DD04QP1&cluster=2235.f. Contact : Feng Liang Email : fliang@lifetech.com
URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DK009DD04QP1.
FEATURES Location/Qualifiers
source 1. .1171
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YH08"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match 7.9%; Score 36; DB 13; Length 1171;
Best Local Similarity 12.1%; Pred. No. 72;
Matches 22; Conservative 95; Mismatches 65; Indels 0; Gaps 0;

```
Qy         71 CTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTG 130
              | | : : ||:::: |:| :  ||:  : : :|::  : :  : ::: ||:::::: :::
Db        947 CCCCKSMGGGSKSMAGMGMKCKCCMGCKCKMKGKSCMKGMRCMGMBMMGCMMMMMKCMKV 1006

Qy        131 AGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCC 190
              |  :  ::: : :: : : :::: :::::::  : :::   : :::: :::|::|:::
Db       1007 KGGCKGGMMKSBGMMTKMMKKKKKMMKKKMMMAKCMMMMMMMMMWKMTMMMTKMTKMMW 1066

Qy        191 GTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGG 250
              :: :  :: ::  :   ::: ::| |::  :::::| : :::   | |  | :|  | ::
Db       1067 KKKMKGYHHDMMMKKKSMKKKMMCKGBMCMKBKBBGCDGMKKMMCCCGCCGCKGCCCCSS 1126

Qy        251 GC 252
              ::
```

Db 1127 SS 1128

RESULT 11
CB642466/c
LOCUS CB642466 675 bp mRNA linear EST 08-APR-2003
DEFINITION OSJNEb02K15.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA clone OSJNEb02K15 3', mRNA sequence.
ACCESSION CB642466
VERSION CB642466.1 GI:29637457
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 675)
AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: gta aaa cga cgg cca gtg
BACKWARD: gga aac agc tat gac cat g
Plate: 02 row: K column: 15
Seq primer: gga aac agc tat gac cat g.
FEATURES Location/Qualifiers
source 1. .675
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEb02K15"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEb"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2: XhoI; 24 hrs after innoculation with Rice Blast (Che 86061)"
ORIGIN

Query Match 7.9%; Score 35.8; DB 14; Length 675;
Best Local Similarity 59.2%; Pred. No. 63;
Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy          6 ACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGT 65
              ||||  | | |  |||||     || ||    ||   | || |||     |||  || |||||
Db        626 ACTTCCACACGGTGAAGCTGCCGCTGCTGCCGCCGTGTCGGGTGAACCTCCTGCGGGTGT 567

Qy         66 TGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGT 108
                ||||  |     |    | ||||    ||||||| |||||||||
Db        566 GCTTCCGGACGGTGTACGTGGCGAGCACCACGGCGGTGGCGGT 524
```

RESULT 12
CNS02APC/c
LOCUS CNS02APC 774 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 252I15 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL188841
VERSION AL188841.1 GI:7826945
KEYWORDS GSS; genome survey sequence.

```
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 774)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="252I15"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG252AE08SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 29;  Length 774;
  Best Local Similarity   48.7%;  Pred. No. 67;
  Matches   97;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
               ||| | | |   |  | || || |  | |  | || |  |  ||| |  | | ||||||
Db         445 GGTGGTGATGGTGGTGATGATGATGATGATGATGGTGATGATGGTGATGATGATGGTGGT 386

Qy         106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
               ||| || || | ||     ||  |||         | |  |  | ||||| |||     |
Db         385 GGTGATGATGGTGGTGATGATGGTGATGATGGTGGTGATGATGATGGTGGTGGTGGTGAT 326

Qy         166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
               |    || |         |||||   ||     |||       |  |     |  |||| ||
Db         325 GATGATGGTGGTGGTGGTGATGATGGTGGTGATGATGGTGGTGGTGGTGATGATGGTGGT 266

Qy         226 GGTAATGGTGCAGATGTTG 244
               | | ||||||  |||| ||
Db         265 GATGATGGTGGTGATGATG 247


RESULT 13
BE572585
LOCUS       BE572585                935 bp    mRNA    linear   EST 15-AUG-2000
DEFINITION  601329541F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3603347 5',
            mRNA sequence.
ACCESSION   BE572585
VERSION     BE572585.1  GI:9816305
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8792  row: a  column: 12
            High quality sequence stop: 679.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3603347"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 10;  Length 935;
  Best Local Similarity   53.1%;  Pred. No. 73;
  Matches   76;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         104 GCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACT 163
               || | ||  | | | | |||||  |    ||  | ||  | | | |    |  ||| ||
Db         485 GCTGAAAAGAGATCTGGCCAAAGACATCACTTCAGATACATCTGGAGACTTTCGGAAAGC 544

Qy         164 CTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCG 223
               ||   ||||||||    |  | || ||||      ||||||| | |  ||  || | |
Db         545 CTTTGCTTGCTCTTGCCAAGGGTGACCGTTGTCAGGACTTGAGTGTGAATCAAGATTGGG 604

Qy         224 GCGGTAATGGTGCAGATGTTGGT 246
               | | ||  | |||     || ||
Db         605 GTGATACAGATGCCAGGCTTTGT 627


RESULT 14
BE604165
LOCUS       BE604165                435 bp    mRNA    linear   EST 21-AUG-2000
DEFINITION  WHE1413-1416_I13_I13ZS Wheat drought stressed leaf cDNA library
            Triticum aestivum cDNA clone WHE1413-1416_I13_I13, mRNA sequence.
ACCESSION   BE604165
VERSION     BE604165.1  GI:9861435
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Han,P.S., Hsia,C.C.,
            Kang,Y., Lazo,G.R., Miller,R., Nguyen H.T., Rausch,C.J.,
            Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
```

```
            genomes - Drought stressed leaf cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="TAM W101"
                     /db_xref="taxon:4565"
                     /clone="WHE1413-1416_I13_I13"
                     /tissue_type="Leaf"
                     /dev_stage="Full tillering stage"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat drought stressed leaf cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were given a gradual
                     stress down to 65% and 78% RWC at Texas  Tech University
                     (D. Zhang in HT Nguyen lab).  Total RNA and poly(A) RNA
                     were prepared, a cDNA library was made, and the cDNA
                     clones were in  vivo excised to give pBluescript phagemids
                     in the TJ Close lab  (Choi, Close), at the University of
                     California, Riverside.   Plasmid DNA preparations and DNA
                     sequencing were performed in the  OD Anderson lab (all
                     other authors)."
ORIGIN

  Query Match             7.8%;  Score 35.6;  DB 10;  Length 435;
  Best Local Similarity   50.0%;  Pred. No. 57;
  Matches   89;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        134 TGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTA 193
              | ||   |  ||   || ||    ||| |   | || |||   ||||| | | ||||   |
Db         10 TCAAATTGGACGTCTATGGAAATGGTGAAGATTCAGTTGAAGTGCAATCAGCTGCCAAGA 69

Qy        194 ACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCT 253
              |    || |||| | ||| |   |||   ||| | ||  |||||||| |  ||
Db         70 AGCTGGATTTGAATCTTAACTTCCATAAAGGCCGAGATCATGCAGATGATTCTCTTCATG 129

Qy        254 CAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGA 311
                | |    |   |  | | || |    |||   ||     ||.||| |||  |||
Db        130 GGTACAAGGTTTTTGTAAACCCGAGCATTAGTGATGTCCTCTGCACCGCCACTGCTGA 187


RESULT 15
BE604467
LOCUS       BE604467               498 bp    mRNA    linear   EST 21-AUG-2000
DEFINITION  WHE1413-1416_B07_B07ZS Wheat drought stressed leaf cDNA library
            Triticum aestivum cDNA clone WHE1413-1416_B07_B07, mRNA sequence.
ACCESSION   BE604467
VERSION     BE604467.1  GI:9861825
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Han,P.S., Hsia,C.C.,
            Kang,Y., Lazo,G.R., Miller,R., Nguyen H.T., Rausch,C.J.,
            Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Drought stressed leaf cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="TAM W101"
                     /db_xref="taxon:4565"
                     /clone="WHE1413-1416_B07_B07"
                     /tissue_type="Leaf"
                     /dev_stage="Full tillering stage"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat drought stressed leaf cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were given a gradual
                     stress down to 65% and 78% RWC at Texas  Tech University
                     (D. Zhang in HT Nguyen lab).  Total RNA and poly(A) RNA
                     were prepared, a cDNA library was made, and the cDNA
                     clones were in  vivo excised to give pBluescript phagemids
                     in the TJ Close lab  (Choi, Close), at the University of
                     California, Riverside.   Plasmid DNA preparations and DNA
                     sequencing were performed in the  OD Anderson lab (all
                     other authors)."
ORIGIN

  Query Match             7.8%;  Score 35.6;  DB 10;  Length 498;
  Best Local Similarity   50.0%;  Pred. No. 61;
  Matches   89;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        134 TGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTA 193
              | ||   |  ||   || ||    ||| |   | || |||   ||||| | | ||||   |
Db         10 TCAAATTGGACGTCTATGGAAATGGTGAAGATTCAGTTGAAGTGCAATCAGCTGCCAAGA 69

Qy        194 ACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCT 253
              |    || |||| | ||| |   |||   ||| | ||  |||||||| |  ||
Db         70 AGCTGGATTTGAATCTTAACTTCCATAAAGGCCGAGATCATGCAGATGATTCTCTTCATG 129

Qy        254 CAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGA 311
                | |    |   |  | | || |    |||   ||     || ||| |||  |||
Db        130 GGTACAAGGTTTTTGTAAACCCGAGCATTAGTGATGTCCTCTGCACCGCCACTGCTGA 187


Search completed: March 18, 2004, 20:11:15
Job time : 2037.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		456	100.0	4680	1	ECCSGABDG	X90754 E.coli csgG
2		456	100.0	10346	1	AE000205	AE000205 Escherich
3		456	100.0	15047	1	D90741	D90741 Escherichia
4		454.4	99.6	1711	1	AF275733	AF275733 Escherich
5		454.4	99.6	10190	1	AE005315	AE005315 Escherich
6		454.4	99.6	327773	1	AP002554	AP002554 Escherich
7		451.2	98.9	10370	1	AE015131	AE015131 Shigella
8		451.2	98.9	306358	1	AE016759	AE016759 Escherich
9		449.6	98.6	292504	1	AE016981	AE016981 Shigella
10		356.8	78.2	2889	1	CSP515700	AJ515700 Citrobact
11		332.8	73.0	2067	1	SEU43280	U43280 Salmonella
12		332.8	73.0	5103	1	STAJ2301	AJ002301 Salmonell
13		332.8	73.0	22411	1	AE008749	AE008749 Salmonell
14		331.2	72.6	254050	1	AL627269	AL627269 Salmonell
15	c	331.2	72.6	301983	1	AE016840	AE016840 Salmonell
16		323	70.8	2920	1	CFR515701	AJ515701 Citrobact
17		275	60.3	1212	1	ECO131756	AJ131756 Escherich
18		266.4	58.4	1048	1	STAGFBA	AJ000514 Salmonell
19		258.2	56.6	2883	1	ESA515702	AJ515702 Enterobac
20		51.2	11.2	301214	1	AE016786	AE016786 Pseudomon
21		49.4	10.8	246435	2	AC096956	AC096956 Rattus no
22	c	47.6	10.4	253001	3	AE014834	AE014834 Plasmodiu
23	c	43.4	9.5	234090	2	AC111458	AC111458 Rattus no
24		43.2	9.5	182148	2	AC119000	AC119000 Rattus no
25		43	9.4	3732	3	PFASA27	M10129 Plasmodium
26		42.6	9.3	210554	2	AC133982	AC133982 Rattus no
27	c	41.4	9.1	10709	1	AE015532	AE015532 Shewanell
28	c	40.4	8.9	172742	2	AC136466	AC136466 Rattus no
29	c	40.2	8.8	242510	2	AC128507	AC128507 Rattus no
30	c	39.8	8.7	160460	2	AC024374	AC024374 Homo sapi
31	c	39.8	8.7	161536	2	AC024318	AC024318 Homo sapi
32		39.8	8.7	179693	9	AL591856	AL591856 Human DNA
33		39.8	8.7	192016	2	AC026540	AC026540 Homo sapi
34	c	39.6	8.7	7218	6	I66494	I66494 Sequence 14
35	c	39.6	8.7	257757	3	AE014837	AE014837 Plasmodiu
36		39.2	8.6	74524	9	AL512508	AL512508 Human DNA
37	c	39.2	8.6	190950	2	AL359639	AL359639 Homo sapi
38	c	38.6	8.5	50820	1	AE014163	AE014163 Streptoco
39		38.6	8.5	310850	1	AP005142	AP005142 Streptoco
40		38.2	8.4	4860	6	AR320671	AR320671 Sequence
41	c	38.2	8.4	186401	2	AC097009	AC097009 Pan trogl
42	c	38.2	8.4	242730	2	AC099408	AC099408 Pan trogl
43	c	37.8	8.3	2000	6	AX655393	AX655393 Sequence
44	c	37.8	8.3	31521	3	CEY47H10A	AL132848 Caenorhab
45	c	37.8	8.3	141489	10	AL591376	AL591376 Mouse DNA

ALIGNMENTS

RESULT 1
ECCSGABDG
LOCUS ECCSGABDG 4680 bp DNA linear BCT 07-JUL-2002
DEFINITION E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION X90754
VERSION X90754.1 GI:1147558
KEYWORDS csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene; orfC gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
TITLE Expression of two csg operons is required for production of

```
          fibronectin- and congo red-binding curli polymers in Escherichia
          coli K-12
  JOURNAL Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE 96414468
   PUBMED 8817489
REFERENCE 2  (bases 1 to 4680)
  AUTHORS Hammar,M.
  TITLE   Direct Submission
  JOURNAL Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
          Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
          SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           100.0%;  Score 456;  DB 1;  Length 4680;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-116;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3233 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 3292

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3293 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 3352

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3353 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 3412

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3413 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 3472

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3473 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 3532

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3533 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 3592

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3593 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 3652

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       3653 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 3688
```

```
RESULT 2
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
```

```
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
protein_bind         1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
protein_bind         1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
promoter             1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
gene                 1494. .1745
                     /gene="b1030"
CDS                  1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
gene                 1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
CDS                  1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
gene                 complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
tRNA                 complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
promoter             2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
promoter             2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
promoter             2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
gene                 2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
CDS                  2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
promoter             3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
promoter             3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
promoter             3396. .3423
```

Query Match 100.0%; Score 456; DB 1; Length 10346;
Best Local Similarity 100.0%; Pred. No. 2.7e-116;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8497 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 8556

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8557 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 8616

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8617 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 8676

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8677 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 8736

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8737 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 8796

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8797 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 8856

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8857 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 8916

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       8917 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 8952
```

```
RESULT 3
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
```

```
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                      Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .15047
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:83333"
                     /map="23.7-24.0 min"
                     /clone="Kohara clone #231"
                     /note="Nucleotide position 1097101-1112147 from the
                     initiation site of ThrA (0 min.).~This clone is from
                     Kohara lambda miniset library."
     CDS             <1. .276
                     /note="unnamed protein product; ORF_ID:o230#5"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35822.1"
                     /db_xref="GI:4062605"
                     /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
                     WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     CDS             1152. .2090
                     /note="ORF_ID:o231#1
                     similar to SwissProt Accession Number P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical 33.0 kd protein in proB-proA
                     intergenic region."
                     /protein_id="BAA35823.1"
                     /db_xref="GI:4062606"
                     /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
                     WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
                     AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
                     PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
                     GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
                     AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     CDS             2145. .2882
                     /note="unnamed protein product; ORF_ID:o231#2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35824.1"
                     /db_xref="GI:4062607"
                     /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                     EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                     PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                     SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
                     ILNVSPRRLLNFLESRGMAPIAEFADL"
     CDS             2906. .3460
                     /note="ORF_ID:o231#3
                     similar to PIR Accession Number F64035"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein HI1543"
                     /protein_id="BAA35825.1"
                     /db_xref="GI:4062608"
                     /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                     DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                     ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                     RTMAPLTRDAISAMWDELEEDSEE"
     CDS             3514. .4053
                     /note="unnamed protein product; ORF_ID:o231#4"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35826.1"
                     /db_xref="GI:4062609"
                     /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                     SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                     IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                     LAARSAKTAHREQEIKNKA"
     gene            complement(4117. .4950)
                     /gene="csgG"
     CDS             complement(4117. .4950)
                     /gene="csgG"
                     /note="ORF_ID:o231#5
                     similar to SwissProt Accession Number P52103"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgG precursor."
                     /protein_id="BAA35827.1"
                     /db_xref="GI:4062610"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(4977. .5393)
                     /gene="csgF"
     CDS             complement(4977. .5393)
                     /gene="csgF"
                     /note="ORF_ID:o231#6
                     similar to SwissProt Accession Number P52104"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgF precursor."
                     /protein_id="BAA35828.1"
                     /db_xref="GI:4062611"
```

```
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(5418. .5807)
                     /gene="csgE"
     CDS             complement(5418. .5807)
                     /gene="csgE"
                     /note="ORF_ID:o231#7
                     similar to SwissProt Accession Number P52105"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgE precursor."
                     /protein_id="BAA35829.1"
                     /db_xref="GI:4062612"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(5812. .6462)
                     /gene="csgD"
     CDS             complement(5812. .6462)
                     /gene="csgD"
                     /note="ORF_ID:o231#8
                     similar to SwissProt Accession Number P52106"
                     /codon_start=1
                     /transl_table=11
                     /product="Probable csgAB operon transcriptional regulatory
                     protein."
                     /protein_id="BAA35830.1"
                     /db_xref="GI:4062613"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            7217. .7672
                     /gene="csgB"
     CDS             7217. .7672

  Query Match           100.0%;  Score 456;  DB 1;  Length 15047;
  Best Local Similarity  100.0%;  Pred. No. 2.6e-116;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7217 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 7276

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7277 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 7336

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7337 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 7396

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7397 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 7456

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7457 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 7516

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7517 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 7576

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7577 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 7636

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       7637 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 7672
```

```
RESULT 4
AF275733
LOCUS       AF275733               1711 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
            CsgA protein (csgA) genes, complete cds.
ACCESSION   AF275733
VERSION     AF275733.1  GI:14039399
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Mutations in the csgD promoter associated with variations in curli
            expression in certain strains of Escherichia coli O157:H7
  JOURNAL   Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
  MEDLINE   21218556
   PUBMED   11319125
REFERENCE   2  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
            Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES             Location/Qualifiers
     source          1. .1711
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="43895 Red Variant"
                     /serotype="O157:H7"
                     /db_xref="taxon:562"
     gene            757. .1212
                     /gene="csgB"
     CDS             757. .1212
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="AAK53211.1"
                     /db_xref="GI:14039400"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1253. .1711
                     /gene="csgA"
     CDS             1253. .1711
                     /gene="csgA"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="AAK53212.1"
                     /db_xref="GI:14039401"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           99.6%;  Score 454.4;  DB 1;  Length 1711;
  Best Local Similarity  99.8%;  Pred. No. 7.6e-116;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 816

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 876

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        877 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 936

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        937 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 996

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        997 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 1056

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1057 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 1116

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 1176

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1177 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 1212


RESULT 5
AE005315
LOCUS       AE005315               10190 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
            of 155.
ACCESSION   AE005315 AE005174
VERSION     AE005315.1  GI:12514572
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10190
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            617. .1072
                     /gene="csgB"
                     /note="synonym: Z1675"
     CDS             617. .1072
                     /gene="csgB"
                     /function="structural component; Cell exterior
                     constituents: Surface structures"
                     /note="Residues 1 to 151 of 151 are 100.00 pct identical
                     to residues 1 to 151 of 151 from Escherichia coli K-12
                     Strain MG1655: B1041"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAG55787.1"
                     /db_xref="GI:12514573"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1113. .1571
                     /gene="csgA"
                     /note="synonym: Z1676"
     CDS             1113. .1571
                     /gene="csgA"
                     /function="regulator; Cell exterior constituents: Surface
                     structures"
                     /note="Residues 1 to 152 of 152 are 96.71 pct identical to
                     residues 1 to 151 of 151 from Escherichia coli K-12 Strain
                     MG1655: B1042"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin major subunit, coiled surface structures;
                     cryptic"
                     /protein_id="AAG55788.1"
                     /db_xref="GI:12514574"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
     gene            1630. .1962
                     /gene="csgC"
                     /note="synonym: Z1677"
     CDS             1630. .1962
                     /gene="csgC"
                     /function="putative structure; Not classified"
                     /note="Residues 1 to 110 of 110 are 98.18 pct identical to
                     residues 1 to 110 of 110 from Escherichia coli K-12 Strain
                     MG1655: B1043"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="AAG55789.1"
                     /db_xref="GI:12514575"
                     /translation="MNALLLLAALSSQITFNTTQQGDMYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     gene            2083. .2394
                     /gene="Z1678"
     CDS             2083. .2394
                     /gene="Z1678"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 103 of 103 are 98.05 pct identical to
                     residues 1 to 103 of 103 from Escherichia coli K-12 Strain
                     MG1655: B1044"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55790.1"
                     /db_xref="GI:12514576"
                     /translation="MFRPFLNSLMLGSLFFPFIAIAGSTAQGGVIHPYGQIVEPACDV
                     STQSSPVEMNCPQNGSVPGKTYSSKALMSGNVKNAQIASVKVQYLDKQKKLAVMNIEY
                     N"
     gene            2489. .3022
                     /gene="Z1679"
     CDS             2489. .3022
                     /gene="Z1679"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 177 of 177 are 100.00 pct identical
                     to residues 1 to 177 of 177 from Escherichia coli K-12
                     Strain MG1655: B1045"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyprotein"
                     /protein_id="AAG55791.1"
```

/db_xref="GI:12514577"
/translation="MKTRIHVVQGDITKLAVDVIVNAANPSLMGGGGVDGAIHRAAGP
ALLDACLKVRQQQGDCPTGHAVITLAGDLPAKAVVHTVGPVWRGGEQNEDQLLQDAYL
NSLRLVAANSYTSVAFPAISTGVYGYPRAAAAEIAVKTVSEFITRHALPEQVYFVCYD
EENAHLYERLLTQQGDE"

gene 2964. .4445
/gene="ymdC"
/note="synonym: Z1680"

CDS 2964. .4445
/gene="ymdC"
/function="putative enzyme; Not classified"
/note="Residues 1 to 493 of 493 are 99.59 pct identical to residues 1 to 493 of 493 from Escherichia coli K-12 Strain MG1655: B1046"
/codon_start=1
/transl_table=11
/product="putative synthase"
/protein_id="AAG55792.1"
/db_xref="GI:12514578"
/translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPGQCGLFPL
EKSLDAFAARYRLAEMSEHTLDVQYYIWQDDMSGRLLFSALLAAAKRGVRVRLLLDDN
NTPGLDDILRLLDSHPRIEVRLFNPFSFRLLRPLGYITDFSRLNRRMHNKSFTVDGVV
TLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWYCKSVSPLQQVLDV
PEGEMADRIELPASWHNDAMTHRYLRKMESSPFINHLVDGTLPLIWAKTRLLSDDPAK
GEGKAKRHSLLPQRLFDIMGSPSERIDIISSYFVPTRAGVAQLLRMVRKXVKIAILTN
SLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREQSSTLHDRGITGNSGASLHAKT
FSIDGKTVFIGSFNFDPRSTLLNTEMGFVIESETLAQLIDKRFIQSQYDAAWQLRLDR
WGRINWVDRHAKKEIVLKKEPATSFWKRVMVRLASILPVEWLL"

gene complement(4453. .5610)
/gene="ymdD"
/note="synonym: Z1681"

CDS complement(4453. .5610)
/gene="ymdD"
/function="orf; Unknown function"
/note="Residues 1 to 385 of 385 are 99.22 pct identical to residues 1 to 385 of 385 from Escherichia coli K-12 Strain MG1655: B1047"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55793.1"
/db_xref="GI:12514579"
/translation="MNPVPAQREYFXDSIRAWLMLLGIPFHISLIYSSHTWHVNSAEP
SLWLTLFNDFIHSFRMQVFFVISGYFSYMLFLRYPLKKWWKVRVERVGIPMLTAIPLL
TLPQFIMLQYVKGKAESWPGLSLYDKYNTLAWELISHLWFLLVLVVMTTLCVWIFKRI
RNNLENSDKTNKKFSMVKLSVIFLCLGIGYAVIRRTIFIVYPPILSNGTFNFIVMQTL
FYLPFFILGALAFIFPHLKALFTTPSRGCTLAAALAFVAYLLNQRYGSGDAWMYETES
VITMVLGLWMVNVVFSFGHRLLNFQSARVTYFVNASLFIYLVHHPLTLFFGAYITPHI
TSNWLGFLCGLIFVVGIAIILYEIHLRIPLLKFLFSGKPVVKRENDKAPAR"

gene 6003. .7538
/gene="mdoG"
/note="synonym: Z1683"

CDS 6003. .7538
/gene="mdoG"
/function="enzyme; Osmotic adaptation"
/note="Residues 1 to 511 of 511 are 99.80 pct identical to residues 1 to 511 of 511 from Escherichia coli K-12 Strain MG1655: B1048"
/codon_start=1
/transl_table=11
/product="periplasmic glucans biosynthesis protein"
/protein_id="AAG55794.1"
/db_xref="GI:12514580"
/translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAQSLAGKGYEAPK
SNLPSVFRDMKYADYQQIQFNHDKAYWNNLKTPFKLEFYHQGMYFDTPVKINEVTATA
VKRIKYSPDYFTFGDVQHDKDTVKDLGFAGFKVLYPINSKDKNDEIVSMLGASYFRVI
GAGQVYGLSARGLAIDTALPSGEEFPRFKEFWIERPKPTDKRLTIYALLDSPRATGAY
KFVVMPGRDTVVDVQSKIYLRDKVGKLGVAPLTSMFLFGPNQPSPANNYRPELHDSNG
LSIHAGNGEWIWRPLNNPKHLAVSSFSMENPQGFGLLQRGRDFSRFEDLDDRYDLRPS
AWVTPKGEWGKGSVELVEIPTNDETNDNIVAYWTPDQLPEPGKEMNFKYTITFSRDED
KLHAPDNAWVQQTRRSTGDVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDTPVTAQTS
IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDAKKTTEMRAALVNADQTLSETWSYQLP
ANE"

gene 7531. .10074
/gene="mdoH"
/note="synonym: Z1684"

CDS 7531. .10074
/gene="mdoH"
/function="enzyme; Osmotic adaptation"

Query Match 99.6%; Score 454.4; DB 1; Length 10190;
Best Local Similarity 99.8%; Pred. No. 7.4e-116;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 676

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 736

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 796

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 856

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 916

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        917 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 976

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 1036

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1037 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 1072
```

RESULT 6
AP002554
LOCUS AP002554 327773 bp DNA linear BCT 21-DEC-2002
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION AP002554 BA000007
VERSION AP002554.1 GI:13360491
KEYWORDS .
SOURCE Escherichia coli O157:H7
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak
JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE 20198780
PUBMED 10734605
REFERENCE 2
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.
TITLE Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655

```
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
   PUBMED   11108008
REFERENCE   3
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
   PUBMED   11111050
REFERENCE   4
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
   PUBMED   11258796
REFERENCE   5  (bases 1 to 327773)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .327773
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            150. .2258
                     /gene="ECs1032"
     CDS             150. .2258
                     /gene="ECs1032"
                     /note="similar to YCBY_ECOLI gi|1787181 percent identity
                     99 in 702 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative oxidoreductase"
                     /protein_id="BAB34455.1"
                     /db_xref="GI:13360492"
                     /translation="MNSLFASTARGLEELLKTELENLGAVECQVVQGGVHFKGDTRLV
                     YQSLMWSRLASRIMLPLGECKVYSDLDLYLGVQAINWTEMFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRDAPDIRVNVWLHKETASIALDLSGDGLH
                     LRGYRDRAGIAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMLATDRAPGL
                     HRGRWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIQRARTNAR
                     LAGIGELITFEVKDVAQLTNPLPKEPYGTVLSNPPYGERLDSEPALIALHSLLGRIMK
                     NQFGGWNLSLFSASPDLLSCLQLRADKQYKAKNGPLDCVQKNYHVAESTPDSKPAMVA
                     EDYANRLRKNLKKFEKWARQEGIECYRLYDADLPEYNVAVDRYADWVVVQEYAPPKTI
                     DAHKARQRLFDIIAATISVLGIAPNKLVLKTRERQKGKNQYQKLGEKGEFLEVTEYNA
                     HLWVNLTDYLDTGLFLDHRIARRMLGQMSKGKDFLNLFSYTGSATVHAGLGGARSTTT
                     VDMSRTYLEWAERNLRLNGLTGRAHRLIQADCLAWLREANEQFDLIFIDPPTFSNSKR
                     MEDAFDVQRDHLALMKDLKRLLRAGGTIMFSNNKRGFRMDLDGLAKLGLKAQEITQKT
                     LSQDFARNRQIHNCWLITAA"
     gene            2270. .4177
                     /gene="ECs1033"
     CDS             2270. .4177
                     /gene="ECs1033"
                     /note="similar to UUP_ECOLI gi|1787182 percent identity 99
                     in 635 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative ATP-binding component of a transport
                     system"
                     /protein_id="BAB34456.1"
                     /db_xref="GI:13360493"
                     /translation="MSLISMHGAWLSFSDAPLLDNAELHIEDNERVCLVGRNGAGKST
                     LMKILNREQGLDDGRIIYEQDLIVARLQQDPPRNVEGSVYDFVAEGIEEQAEYLKRYH
                     DISRLVMNDPSEKNLNELAKVQEQLDHHNLWQLENRINEVLAQLGLDPNVALSSLSGG
                     WLRKAALGRALVSNPRVLLLDEPTNHLDIETIDWLEGFLKTFNGTIIFISHDRSFIRN
                     MATRIVDLDRGKLVTYPGNYDQYLLEKEEALRVEELQNAEFDRKLAQEEVWIRQGIKA
                     RRTRNEGRVRALKAMRRERGERREVMGTAKMQVEEASRSGKIVFEMEDVCYQVDGKQL
                     VKDFSAQVLRGDKIALIGPNGCGKTTLLKLMLGQLQADSGRIHVGTKLEVAYFDQHRA
                     ELDPDKTVMDNLAEGKQEVMVNGKPRHVLGYLQDFLFHPKRAMTPVRALSGGERNRLL
                     LARLFLKPSNLLILDEPTNDLDVETLELLEELIDSYQGTVLLVSHDRQFVDNTVTECW
                     IFEGGGKIGRYVGGYHDARGQQEQYVALKQPAVKKNEEPAAPKAETVKRSSSKLSYKL
                     QRELEQLPQLLEDLEAKLEALQTQVADASFFSQPHEQTQKVLADMAAAEQELEQAFER
                     WEYLEALKNGG"
     gene            4307. .5560
                     /gene="ECs1034"
     CDS             4307. .5560
                     /gene="ECs1034"
                     /note="similar to PQIA_ECOLI gi|1787183 percent identity
                     99 in 417 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="paraquat-inducible protein A"
                     /protein_id="BAB34457.1"
                     /db_xref="GI:13360494"
                     /translation="MCEHHHAAKHILCSQCDMLVALPRLEHGQKAACPRCGTTLTVAW
                     DAPRQRPTAYALAALFMLLLSNLFPFVNMNVAGVTSEITLLEIPGVLFSEDYASLGTF
                     FLLFVQLVPAFCLITILLLVNRAELPVRLKEQLARVLFQLKTWGMAEIFLAGVLVSFV
                     KLMAYGSIGVGSSFLPWCLFCVLQLRAFQCVDRRWLWDDIAPMPELRQPLKPGVTGIR
                     QGLRSCSCCTAILPADEPVCPRCGTKGYVRRRNSLQWTLALLVTSIMLYLPANILPIM
                     VTDLLGSKMPSTILAGVILLWSEGSYPVAAVIFLASIMVPTLKMIAIAWLCWDAKGHG
                     KRDSERMHLIYEVVEFVGRWSMIDVFVIAVLSALVRMGGLMSIYPAMGALMFALVVIM
                     TMFSAMTFDPRLSWDRQPESEHEES"
     gene            5565. .7205
                     /gene="ECs1035"
     CDS             5565. .7205
                     /gene="ECs1035"
                     /note="similar to PQIB_ECOLI gi|1787184 percent identity
                     99 in 546 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="paraquat-inducible protein B"
                     /protein_id="BAB34458.1"
                     /db_xref="GI:13360495"
                     /translation="MESNNGEAKIQKVKNWSPVWIFPIVTALIGAWVLFYHYSHQGPE
                     VTLITANAEGIEGGKTTIKSRSVDVGVVESATLADDLTHVEIKARLNSGMEKLLHKDT
                     VFWVVKPQIGREGISGLGTLLSGVYIELQPGAKGSKMDKYDLLDSPPLAPPDAKGIRV
                     VLDSKKAGQLSPGDPVLFRGYRVGSVETSTFDTQKRNISYQLFINAPYDRLVTSNVRF
                     WKDSGIAVDLTSAGMRVEMGSLTTLLSGGVSFDVPEGLDLGQPVAPKTAPVLYDDQKS
                     IQDSLYTDHIDYLMFFKDSVRGLQPGAPVEFRGIRLGTVSKVPFFAPNMRQTFNDDYR
                     IPVLIRIEPERLKMQLGENADVVEHLGELLKRGLRGSLKTGNLVTGALYVDLDFYPNT
                     PAITGIREFNGYQIIPTVSGGLAQIQQRLMEALDKINKLPLNPMIEQATSTLSESQRT
                     MKNLQTTLDSMNKILASQSMQQLPTDMQSTLRELNRSMQGFQPGSAAYNKMVADMQRL
                     DQVLRELQPVLKTLNEKSNALVFEAKDKKDPEPKRAKQ"
     gene            7217. .7765
                     /gene="ECs1036"
     CDS             7217. .7765
                     /gene="ECs1036"
                     /note="similar to YMBA_ECOLI gi|1787185 percent identity
                     98 in 182 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB34459.1"
                     /db_xref="GI:13360496"
```

```
                     /translation="MTIAALWLAGCSSGEINKNYYQLPVVQSGTQSTASQGNRLLWVE
                     QVAVPDYLAGNGVVYQTSDVKYVIANNNLWASPLDQQLRNTLVANLSTQLPGWVVASQ
                     PLGSAQDTLNVTVTEFNGRYDGKVIVSGEWLLNHQGQLIKRPFRLEGVQTQDGYDEMV
                     KVLAGVWSQEAASIAQEIKRLP"
     gene            8021. .8188
                     /gene="ECs1037"
     CDS             8021. .8188
                     /gene="ECs1037"
                     /note="identical to RMF_ECOLI gi|1787186 (Conserved in
                     E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ribosome modulation factor"
                     /protein_id="BAB34460.1"
                     /db_xref="GI:13360497"
                     /translation="MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGW
                     REAMADRVVMA"
     gene            complement(8258. .8776)
                     /gene="ECs1038"
     CDS             complement(8258. .8776)
                     /gene="ECs1038"
                     /note="similar to FABA_ECOLI gi|1787187 percent identity
                     100 in 172 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11

  Query Match             99.6%;  Score 454.4;  DB 1;  Length 327773;
  Best Local Similarity   99.8%;  Pred. No. 7e-116;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325925 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 325984

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325985 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 326044

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326045 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 326104

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326105 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 326164

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326165 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 326224

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326225 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 326284

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326285 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 326344

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db     326345 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 326380


RESULT 7
AE015131
LOCUS       AE015131               10370 bp    DNA     linear   BCT 18-OCT-2002
DEFINITION  Shigella flexneri 2a str. 301 section 94 of 412 of the complete
            genome.
ACCESSION   AE015131 AE005674
VERSION     AE015131.1  GI:24051313
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 301
  ORGANISM  Shigella flexneri 2a str. 301
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 10370)
  AUTHORS   Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
            Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
            Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
            Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
            Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
  TITLE     Genome sequence of Shigella flexneri 2a: insights into
            pathogenicity through comparison with genomes of Escherichia coli
            K12 and O157
  JOURNAL   Nucleic Acids Res. 30 (20), 4432-4441 (2002)
  PUBMED    12384590
REFERENCE   2  (bases 1 to 10370)
  AUTHORS   Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B.,
            Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y.,
            Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y.,
            Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
            of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.
            China
FEATURES             Location/Qualifiers
     source          1. .10370
                     /organism="Shigella flexneri 2a str. 301"
                     /mol_type="genomic DNA"
                     /strain="301"
                     /serotype="2a"
                     /db_xref="taxon:198214"
     tRNA            complement(127. .214)
                     /product="tRNA-Ser"
                     /note="anticodon: GGA"
     gene            410. .1387
                     /gene="ycdW"
                     /locus_tag="SF1027"
                     /pseudo
     CDS             410. .1387
                     /gene="ycdW"
                     /locus_tag="SF1027"
                     /note="similar to Escherichia coli K12 putative
                     dehydrogenase gi: 1787270 (326 aa). BLAST with identity of
                     99% in 326 aa. This CDS contains an in-frame stop codon.
                     The sequence has been checked and is believed to be
                     correct."
                     /pseudo
                     /codon_start=1
                     /transl_table=11
     gene            1442. .2179
                     /gene="ycdX"
                     /locus_tag="SF1028"
     CDS             1442. .2179
                     /gene="ycdX"
                     /locus_tag="SF1028"
                     /note="Residues 1 to 245 of 245 are 99 pct identical to
                     residues 1 to 245 of a 245 aa protein from Escherichia
                     coli O157:H7 EDL933 ref: NP_287168.1"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, conserved hypothetical protein"
                     /protein_id="AAN42652.1"
                     /db_xref="GI:24051314"
                     /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                     EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                     PVFAPHDKATNTQAMISTIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                     SFLHSRKGSEDNCRAVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPLER
                     ILNVSPRRLLNFLESRGMAPIAEFADL"
     gene            2203. .2757
                     /gene="ycdY"
                     /locus_tag="SF1029"
     CDS             2203. .2757
                     /gene="ycdY"
```

/locus_tag="SF1029"
/note="Residues 1 to 184 of 184 are 99 pct identical to residues 1 to 184 of a 184 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287169.1"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase component"
/protein_id="AAN42653.1"
/db_xref="GI:24051315"
/translation="MNEPSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQDELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEDATEAEVRAFLSERGMLLADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFWRTMAPLTRDAISAMWDELEEDSEE"

gene 2811. .3350
/gene="ycdZ"
/locus_tag="SF1030"

CDS 2811. .3350
/gene="ycdZ"
/locus_tag="SF1030"
/note="Residues 1 to 179 of 179 are 98 pct identical to residues 1 to 179 of a 179 aa protein from Escherichia coli K12 ref: NP_415554.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42654.1"
/db_xref="GI:24051316"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLLSWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITGIVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLVFGYAMKNSGLWLAARSAKTAHREQEIKNKA"

gene complement(3414. .4189)
/gene="csgG"
/locus_tag="SF1031"
/pseudo

CDS complement(3414. .4189)
/gene="csgG"
/locus_tag="SF1031"
/note="similar to Escherichia coli K12 curli production assembly-transport component, 2nd curli operon gi: 1787274 (278 aa). BLAST with identity of 97% in 257 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11

gene complement(4454. .5662)
/locus_tag="SF1032"

CDS complement(4454. .5662)
/locus_tag="SF1032"
/note="Residues 1 to 402 of 402 are 100 pct identical to residues 1 to 402 of a 402 aa protein from Plasmid R100 ref: NP_052925.1"
/codon_start=1
/transl_table=11
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translation="MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTLTLTELGRNLPTKARTKHNIKRIDRLLGNRHLHKERLAVYRWHASFICSGNTMPIVLVDWSDIREQKRLMVLRASVALHGRSVTLYEKAFPLSEQCSKKAHDQFLADLASILPSNTTPLIVSDAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAENWKPISNLHDMSSSHSKTLGYKRLTKSNPISCQILLYKSRSKGRKNQRSTRTHCHHPSPKIYSASAKEPWILATNLPVEIRTPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWLAGVHAQKQGWDKHFQANTVRNRNVLSTVRLGMEVLRHSGYTITREDSLVAATLLTQNLFTHGYVLGKL"

gene complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"

CDS complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"
/note="Residues 1 to 129 of 129 are 98 pct identical to residues 1 to 129 of a 129 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287173.1"
/codon_start=1
/transl_table=11
/product="curli production assembly-transport component, 2nd curli operon"
/protein_id="AAN42656.1"
/db_xref="GI:24051318"
/translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFYRAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLPPLKRDSEKTVVFALIQTEEALNRRQINQALLSTDDLAHDEF"

gene complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"

CDS complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"
/note="Residues 11 to 226 of 226 are 100 pct identical to residues 1 to 216 of a 216 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=11
/product="putative 2-component transcriptional regulator for 2nd curli operon"
/protein_id="AAN42657.1"
/db_xref="GI:24051319"
/translation="MRSIKKWGFIMFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQRSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWPHINGVFYAMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHREKEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"

gene 7766. .8248
/gene="csgB"
/locus_tag="SF1035"

CDS 7766. .8248
/gene="csgB"
/locus_tag="SF1035"
/note="Residues 10 to 160 of 160 are 99 pct identical to residues 1 to 151 of a 151 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287175.1"
/codon_start=1
/transl_table=11
/product="minor curlin subunit precursor, similar ro CsgA"
/protein_id="AAN42658.1"
/db_xref="GI:24051320"
/translation="MYDQVQGDNMKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDASISQGAYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR

Query Match 98.9%; Score 451.2; DB 1; Length 10370;
Best Local Similarity 99.3%; Pred. No. 5.8e-115;
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
Db 7793 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 7852

Qy 61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
Db 7853 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 7912

Qy 121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
Db 7913 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 7972

Qy 181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
Db 7973 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAATCGGGCA 8032

Qy 241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
Db 8033 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 8092

```
Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8093 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 8152

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       8153 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAGTTGTAGTGCAGAGA 8212

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||| ||||||||||||||||||||||
Db       8213 CAGTCGCAAATGGCAATTCGCGTGACACAACGTTAA 8248
```

```
RESULT 8
AE016759
LOCUS       AE016759               306358 bp    DNA     linear   BCT 24-DEC-2002
DEFINITION  Escherichia coli CFT073 section 5 of 18 of the complete genome.
ACCESSION   AE016759 AE014075
VERSION     AE016759.1  GI:26107527
KEYWORDS    .
SOURCE      Escherichia coli CFT073
  ORGANISM  Escherichia coli CFT073
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Extensive Mosiac Structure Revealed by the Complete Genome Sequence
            of Uropathogenic Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
   PUBMED   12471157
REFERENCE   2  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .306358
                     /organism="Escherichia coli CFT073"
                     /mol_type="genomic DNA"
                     /strain="CFT073"
                     /db_xref="taxon:199310"
     gene            complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
     CDS             complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
                     /note="Residues 27 to 1231 of 1245 are 80.16 pct identical
                     to residues  1 to 1209 of 1218 from GenPept.129 :
                     >emb|CAD05883.1| (AL627276) putative ABC transporter
                     protein [Salmonella enterica subsp. enterica serovar
                     Typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP binding cassette (ABC) transporter homolog"
                     /protein_id="AAN79727.1"
                     /db_xref="GI:26107528"
                     /translation="MRDRSVLAFLRNDPASGSYFTDGGGLMPANHTPTPAQSWIVRLA
                     RVCWERKKLSVIVVVASVSTILLAALTPLLTRQAVNDALAGNPARLPWLACGLLLIAF
                     FDFIGNYVRRGYAGMLSLWVQHTLRGRVFDSIQKLDGAGQDALRTGQVISRTNSDLQQ
                     VHTLLQMCPVPLAVFTYYIAGIAVMLWMSPAMTLIVVCVLVCLAITALRARRRVFAQT
                     GMASDQLANLTEHIREVLAQISVVKSCVAEMRETHWLDRQSRQIVRVRIGAVISQAMP
                     GATMLALPVLGQIVLLCYGGWSVMHGRIDLGTFVAFASFLAMLTGPTRVLASFLVIAQ
                     RTQASVERVFALIDTRSQMEDGTESINSQVVGLELENMSFDYHHGDRHILSDISFSLR
                     AGETVAVVGASGSGKSTLLMLLARFYDPCSGKIWLNTSEGRQNLRDIRLEALRRRVGI
                     VFEDAFLFAGTVAENIAYGHPQATADDIRRAAAAAGASDFINALPKGFDSLLTERGTN
                     LSGGQRQRIALARALITAPDVLILDDTTSAVDAVTEAEINTALGRYADEGHMLLVIAR
                     RRSTLQLASRVVVLDKGRMVDTGTPAELEARCPAFRALMTGDSDFLATSHNSHNELWP
                     AEPATQDDVTDTGDKGFVARMTRVPENAVQQALAGKGRKVTSLLKPVAWMFVIAALLI
                     ALDSAAGVGVLILLQHGIDSGVAAGDMSTIGLCALLALCLVIVGWCSYSLQTVFAARA
                     AESVQHSVRLRSFGHMLRLGLPWHEKHADSRLTRMTVDVDSLARFLQNGLAGAATSLV
                     TMFAIAATMFWLDPILALTALSAVPVAALATMIYRRLSTPAYAQARLEIGKVNSTLQE
                     KVSGLRVVQSHGQQELEGARLRALSERFRATRVRAQKYLAVYFPFLTFCTEASYAAVL
                     LVGASQVAAGEMTAGVLAAFFLLLGQFYGPVQQLSGIVDAWQQATASGKHIDELLATE
                     GTENLGSSSVLPVTGALHLDEVTFSYPDSHEPALNKLTLTIPEGMVVAVVGRSGAGKS
                     TLIKLIAGLYFPTHGNIRIGVQMLDDASLTEYRRQIGLVDQDVALFSSDIAENIRYSR
                     PSATNEDVEIASQRAGLYEMVCNLPQGFRTPVNNGGADLSAGQRQLIALARAQLANAH
                     ILLLDEATSCLDRTSEERLMSSLTDVVHAGKHSALIVAHRLTTAQRCDLIAVIDKGLL
                     AEYGTHEQLLSAGGLYTRLWHDSVSSTALHRQHNMKEETPG"
     gene            complement(3882. .5045)
                     /gene="iroB"
                     /locus_tag="c1254"
     CDS             complement(3882. .5045)
                     /gene="iroB"
                     /locus_tag="c1254"
                     /note="Escherichia coli O157:H7 ortholog: z1190"
                     /codon_start=1
                     /transl_table=11
                     /product="Putative glucosyltransferase"
                     /protein_id="AAN79728.1"
                     /db_xref="GI:26107529"
                     /translation="MRRLPDLDRQAERDFLMRILFVGPPLYGLLYPVLSLAQAFRVNG
                     HEVLIASGGQFAQKAAEAGLVVFDAAPGLDSEAGYRHHEAQRKKSNIGTQMGNFSFFS
                     EEMADHLVEFAGHWRPDLIIYPPLGVIGPLIAAKYDIPVVMQTVGFGHTPWHIRGVTR
                     SLTDAYRRHNVGATPRDMAWIDVTPPSMSILENDGEPIIPMQYVPYNGGAVWEPWWER
                     RPDRKRLLVSLGTVKPMVDGLDLIAWVMDSASEVDAEIILHISANARSDLRSLPSNVR
                     LVDWIPMGVFLNGADGFIHHGGAGNTLTALHAGIPQIVFGQGADRPVNARVVAERGCG
                     IIPGDVGLSSNMINAFLNNRSLRKASEEVAAEMAAQPCPGEVAKSLITMVQKG"
     gene            5535. .5834
                     /locus_tag="c1255"
     CDS             5535. .5834
                     /locus_tag="c1255"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79729.1"
                     /db_xref="GI:26107530"
                     /translation="MKPTLLQTVENICLQILFVLFLFTGTLIDVEALVSCSSVAKNVQ
                     RTPASSVSSSVYSASVSRTAADIMLLSNNPLFAPGQVSFFGCSLAASGIHVTILE"
     repeat_region   5867. .6577
                     /insertion_seq="IS1541A-like"
     gene            5961. .6470
                     /locus_tag="c1256"
     CDS             5961. .6470
                     /locus_tag="c1256"
                     /note="Residues 18 to 169 of 169 are 94.07 pct identical
                     to residues  1 to 152 of 152 from SwissProt.40 :
                     >sp|Q57334|T200_SALTY Transposase for insertion sequence
                     element IS200"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79730.1"
                     /db_xref="GI:26107531"
                     /translation="MRSGNCKCQTRNQKGVPMGNEKSLAHTRWNCKYHIVFAPKYRRQ
                     VFYREKRRAIGCILRKLCEWKSVRILEAECCADHIHMLVEIPPKMSVSGFMGYLKGKS
                     SLMPYEQFGDLKFKYRNREFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPY
                     PGSPFTGRK"
     repeat_region   complement(6578. .6940)
                     /note="incomplete"
                     /insertion_seq="IS4"
     gene            complement(6598. .6960)
                     /locus_tag="c1257"
     CDS             complement(6598. .6960)
                     /locus_tag="c1257"
                     /function="IS, phage, Tn"
                     /note="Residues 8 to 120 of 120 are 98.23 pct identical to
                     residues 316 to 428 of 442 from MG1655 : b4278"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="Putative conserved protein"
                     /protein_id="AAN79731.1"
                     /db_xref="GI:26107532"
                     /translation="MPRQPDQGYREIKQTMQLSRLTLRSKKPELVEQELWGVLLAYNL
                     VRYQMIKMAEHLKGYWPNQLSFSESCGMVMRMLMTLQGASPGRIPELMRDLASMGQLV
                     KLPTRRETAFPRVVKERP"
     gene            7651. .8013
                     /locus_tag="c1258"
     CDS             7651. .8013
                     /locus_tag="c1258"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79732.1"
                     /db_xref="GI:26107533"
                     /translation="MLGACFYSWGTSFYYGFPWWGAELVLMMWPEACFMLLQSSLYSL
                     LDGEQVLFSFWELNKKQYEGFEFYQIFSGNIVASCSSCSGVSSNSSAHYARCTDFMRH
                     YSLNNYAACQLGRQACLS"
     gene            7835. .8269
                     /locus_tag="c1259"
     CDS             7835. .8269
                     /locus_tag="c1259"
                     /note="Residues 1 to 139 of 144 are 50.71 pct identical to
                     residues  104 to 242 of 258 from GenPept.129 :
                     >gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
                     [Shigella flexneri]"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79733.1"
                     /db_xref="GI:26107534"
                     /translation="MKDLSFIRFFLAILLLLVPPALEFPVIHRHITPGALTLCVITAL
                     IITLLASLGGRLVSVKCISEISFIRRHRIECMMAGFMIYFWTFSLIAGWYKPQFKKEY
                     QTLHYENARYYVLARYDGRLVLSQSYRSGSRKFVIINGGHRL"
     gene            8477. .8553
                     /locus_tag="c5534"
     tRNA            8477. .8553
                     /locus_tag="c5534"
                     /product="tRNA-Arg"
                     /function="RNA; tRNA"
     gene            complement(8832. .9290)
                     /locus_tag="c1260"
     CDS             complement(8832. .9290)
                     /locus_tag="c1260"
                     /function="putative enzyme; Transposases Insertion
                     Sequence Associated"
                     /note="Escherichia coli O157:H7 ortholog: z4324"
                     /codon_start=1
                     /transl_table=11
                     /product="Putative Transposase"
                     /protein_id="AAN79734.1"
                     /db_xref="GI:26107535"
                     /translation="MASAADARRIVSHYERRWLIEEYHKAWKSGGTCVESLRMQTRDN
                     LERMVVIKAFIAVRMLGLRQEGISEETQNDSCKKILTPTEWKLLWVKLEGKQLPSQTP
                     TLKWACLKLGRWHDSKRTGRPGWVVMWDGWFRLQDMVEGYPVMKSLDQEI"
     gene            complement(9302. .9667)
                     /locus_tag="c1261"
     CDS             complement(9302. .9667)
                     /locus_tag="c1261"
                     /note="Residues 1 to 35 of 121 are 97.14 pct identical to
                     residues  59 to 93 of 316 from GenPept.129 :
                     >gb|AAK18437.1|AF348706_126 (AF348706) putative
                     transposase [Shigella flexneri]"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79735.1"
                     /db_xref="GI:26107536"
                     /translation="MAERLGEIQKRVITVCDREADIWHYLHYKVSHGQRFVVRTAQNS

Query Match             98.9%;  Score 451.2;  DB 1;  Length 306358;
Best Local Similarity   99.3%;  Pred. No. 5.5e-115;
Matches  453;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44773 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 44832

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      44833 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTTGCGGTAAATGAATTGAGTAAG 44892

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44893 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 44952

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      44953 CGGCAGGGAGGCTCAAAACTTTTGACGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 45012

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45013 AAGATTGATCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 45072

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45073 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 45132

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45133 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 45192

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      45193 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 45228
```

```
RESULT 9
AE016981
LOCUS       AE016981               292504 bp    DNA     linear   BCT 22-APR-2003
DEFINITION  Shigella flexneri 2a str. 2457T section 4 of 16 of the complete
            genome.
ACCESSION   AE016981 AE014073
VERSION     AE016981.1  GI:30040616
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 2457T
  ORGANISM  Shigella flexneri 2a str. 2457T
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Complete Genome Sequence and Comparative Genomics of Shigella
            flexneri Serotype 2a Strain 2457T
  JOURNAL   Infect. Immun. 71 (5), 2775-2786 (2003)
   PUBMED   12704152
REFERENCE   2  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .292504
                     /organism="Shigella flexneri 2a str. 2457T"
                     /mol_type="genomic DNA"
                     /strain="2457T"
                     /serotype="2a"
                     /db_xref="taxon:198215"
     gene            complement(223. .441)
                     /gene="infA"
                     /locus_tag="S0884"
```

```
CDS             complement(223. .441)
                /gene="infA"
                /locus_tag="S0884"
                /function="factor; Proteins - translation and
                modification"
                /note="residues 1 to 72 of 72 are 100.00 pct identical to
                residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
                /codon_start=1
                /transl_table=11
                /product="protein chain initiation factor IF-1"
                /protein_id="AAP16348.1"
                /db_xref="GI:30040617"
                /translation="MAKEDNIEMQGTVLETLPNTMFRVELENGHVVTAHISGKMRKNY
                IRILTGDKVTVELTPYDLSKGRIVFRSR"
gene            complement(726. .1430)
                /gene="aat"
                /locus_tag="S0885"
CDS             complement(726. .1430)
                /gene="aat"
                /locus_tag="S0885"
                /function="enzyme; Aminoacyl tRNA synthetases, tRNA
                modification"
                /note="residues 1 to 234 of 234 are 100.00 pct identical
                to residues 1 to 234 of 234 from Escherichia coli K-12 :
                B0885"
                /codon_start=1
                /transl_table=11
                /product="leucyl, phenylalanyl-tRNA-protein transferase"
                /protein_id="AAP16349.1"
                /db_xref="GI:30040618"
                /translation="MRLVQLSRHSIAFPSPEGALREPNGLLALGGDLSPARLLMAYQR
                GIFPWFSPGDPILWWSPDPRAVLWPESLHISRSMKRFHKRSPYRVTMNYAFGQVIEGC
                ASDREEGTWITRGVVEAYHRLHELGHAHSIEVWREDELVGGMYGVAQGTLFCGESMFS
                RMENASKTALLVFCEEFIGHGGKLIDCQVLNDHTASLGACEIPRRDYLNYLNQMRLGR
                LPNNFWVPRCLFSPQE"
gene            complement(1472. .3193)
                /gene="cydC"
                /locus_tag="S0886"
CDS             complement(1472. .3193)
                /gene="cydC"
                /locus_tag="S0886"
                /function="transport; Transport of small molecules: Other"
                /note="residues 1 to 573 of 573 are 97.03 pct identical to
                residues 1 to 573 of 573 from Escherichia coli K-12 :
                B0886"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding component of cytochrome-related
                transport"
                /protein_id="AAP16350.1"
                /db_xref="GI:30040619"
                /translation="MRALLPYLALYKRHKWMLSLGIVLAIVTLLASIGLLTLSGWFLS
                ASAVAGVAGLYSFNYMLPAAGVRGAAITRTAGRYFERLVSHDATFRVLQHLRIYTFSK
                LLPLSPAGLARYRQGELLNRVVADVDTLDHLYLRVISPLVGAFVVIMVVTIGLSFLDF
                TLAFTLGSVMLLTLFLMPPLFYRAGKSTGQNLTHLRGQYRQQLTAWLQGQAELTIFGA
                SDRYRTQLENTEIQWLEAQRRQSELTALSQAIMLLIGALAVILMLWMASGGVGGNAQP
                GALIALFVFCALAAFEALAPVTGAFQHLGQVIASAVRISDLTDQKPEVTFPDTQTRVA
                DRVSLTLRDVQFTYPEQSQQALKGISLQVNAGEHIAILGRTGCGKSTLLQLLTRAWDP
                QQGEILLNDSPIASLNEAALRQTISVVPQRVHLFSATLRDNLLLASPGSSDEALSEIL
                RRVGLEKLLEDAGLNSWLGEGGRQLSGGELRRLAIARTLLHDAPLVLLDEPTEGLDAT
                TESQILELLAKMMREKTVLMVTHRLRGLSRFQQIIVMDNGQIIEQGTHAELLARQGRY
                YQFKQGL"
gene            complement(3194. .4960)
                /gene="cydD"
                /locus_tag="S0887"
CDS             complement(3194. .4960)
                /gene="cydD"
                /locus_tag="S0887"
                /function="transport; Not classified"
                /note="residues 1 to 588 of 588 are 92.85 pct identical to
                residues 1 to 588 of 588 from Escherichia coli K-12 :
                B0887"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding component of cytochrome-related
                transport"
                /protein_id="AAP16351.1"
                /db_xref="GI:30040620"
                /translation="MNKSRQKELTRWLKQQSVISQRWLNISRLLGFVSGILIIAQAWF
                MARILQHMIMENIPREALLLPFTLLVLTFVLRAWVVWLRERVGYHAGQHIRFAIRRQV
                LDRLQQAGPAWIQGKPAGSWATLVLEQIDDMHDYYARYLPQMALAVSVPLLIVVAIFP
                SNWAAALILLGTAPLIPLFMALVGMGAADANRRNFLALARLSGHFLDRLRGMETLRIF
                GRGEAEIESIRSASEDFRQRTMEVLRLAFLSSGILEFFTSLSIALVAVYFGFSYLGEL
                DFGHYDTGVTLAAGFLALILAPEFFQPLRDLGTFYHAKAQAVGAADSLKTFMETPLAH
                PQRGEAELALTDPLTIEAEDLFITSPEGKTLAGPLNFTLPAGQRAVLVGRSGSGKSSL
                LNALSGFLSYQGSLRINGIELRDLSPESWRKHLSWVGQNPQLPAAKLRDNVLLARPDA
                SEQELQAALDNAWVSEFLPLLPQGVDTPVGDQAARLSVGQAQRVAVARALLNPCSLLL
                LDEPAASLDAHSEQRVMEALNAASLRQTTLMVTHQLEDLADWDVIWVMQDGRIIEQGR
                YAELSVAGGPFATLLAHRQEEI"
gene            complement(5083. .6048)
                /gene="trxB"
                /locus_tag="S0888"
CDS             complement(5083. .6048)
                /gene="trxB"
                /locus_tag="S0888"
                /function="enzyme; 2-Deoxyribonucleotide metabolism"
                /note="residues 1 to 321 of 321 are 95.95 pct identical to
                residues 1 to 321 of 321 from Escherichia coli K-12 :
                B0888"
                /codon_start=1
                /transl_table=11
                /product="thioredoxin reductase"
                /protein_id="AAP16352.1"
                /db_xref="GI:30040621"
                /translation="MGTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQL
                TTTTEVENWPGDPNDLTGPLLMARMHEHATKFETEIIFDHINKVDLQNRPFRLNGDNG
                EYTCDALIIATGASARYLGLPSEEAFKGRGVSACATCDGFFYRNQKVAVIGGGNTAVE
                EALYLSNIASEVHLIHRRDGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGV
                TGVRLRDTQNSDNIESLDVAGLFVAIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQ
                TSIPGVFAAGDVMDHIYRQAITSAGTGCMAALDAERYLDGLADAK"
gene            6592. .7086
                /gene="lrp"
                /locus_tag="S0889"
CDS             6592. .7086
                /gene="lrp"
                /locus_tag="S0889"
                /function="regulator; Central intermediary metabolism:
                Pool, multipurpose conversions"
                /note="residues 1 to 164 of 164 are 100.00 pct identical
                to residues 1 to 164 of 164 from Escherichia coli K-12 :
                B0889"
                /codon_start=1
                /transl_table=11
                /product="regulator for leucine (or lrp) regulon and
                high-affinity branched-chain amino acid transport system"
                /protein_id="AAP16353.1"
                /db_xref="GI:30040622"
                /translation="MVDSKKRPGKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTP
                CLERVRRLERQGFIQGYTALLNPHYLDASLLVFVEITLNRGAPDVFEQFNTAVQKLEE
                IQECHLVSGDFDYLLKTRVPDMSAYRKLLGETLLRLPGVNDTRTYVVMEEVKQSNRLV
                IKTR"
gene            7221. .11249
                /gene="ftsK"
                /locus_tag="S0890"
CDS             7221. .11249
                /gene="ftsK"
                /locus_tag="S0890"
                /function="phenotype; Cell division"
                /note="residues 1 to 737 of 1342 are 68.24 pct identical
                to residues 1 to 737 of 1329 from Escherichia coli K-12 :
                B0890"
                /codon_start=1
                /transl_table=11
                /product="cell division protein"
                /protein_id="AAP16354.1"
                /db_xref="GI:30040623"
```

```
                     /translation="MSQEYTEDKEVTLTKLSSGRRLLEALLILIVLFAVWLMAALLSF
                     NPSDPSWSQTAWHEPIHNLGGMPGAWLADTLFFIFGVMAYTIPVIIVGGCWFAWRHQS
                     SDEYIDYFAVSLRIIGVLALILTSCGLAAINADDIWYFASGGVIGSLLSTTLQPLLHS
                     SGGTIALLCVWAAGLTLFTGWSWVTIAEKLGGWILNILTFASNRTRRDDTWVDEDEYE
                     DDEEYEDENHGKQHESRRARILRGALARRKRLAEKFINPMGRQTDAALFSGKRMDDDE
                     EITYTARGVAADPDDVLFSGNRATQPEYDEYDPLLNSAPITEPVAVAAAATTATQSWA
                     APVEPVTQTPPVASVDVPPSQPTVAWQPVPGPQTGEPVIAPAPEGYPQQSQYAQPAVQ
                     YNEPLQQPVQPQQPYYAPAAEQPAQQPYYAPAPEQPVAGNAWQAEEQQSTFAPLSTYQ
                     TEQTYQQPAAQEPLYQQPQPVEQQPVVEPEPVVEETKPARPPLYYFEEVEEKRARERE
                     QLAAWYQPIPEPVKEPEPIKSSLKAPSVAAVPPVEAAAAAVSPLASGVKKATLATGAAA
                     TVAAPVFSLANSGGPRPQVKEGIGPQLPRPKRIRVPTRRELASYGIKLPSQRAAEEKA
                     REAQRNQYDSGDQYNDDEIDAMQQDELARQFAQTQQQRYGEQYQHDVPVNAEDADAAA
                     EAELARQFAQTQQQRYSGEQPAGANPFSLDDFEFSPMKALLDDGPHEPLFTPIVEPVQ
                     QPQQPVAPQQQYQQPQQPVAPQQQYQQPQQQVAPQPQYQQPQQPVAPQQQYQQPQQPV
                     APQPQYQQPQQPVAPQPQYQQPQQPVAPQPQDTLLHPLLMRNGDSRPLHKPTTPLPSL
                     DLLTPPPSEVEPVDTFALEQMARLVEARLADFRIKADVVNYSPGPVITRFELNLAPGV
                     KAARISNLSRDLARSLSTVAVRVVEVIPGKPYVGLELPNKKRQTVYLREVLDNAKFRD
```

Query Match 98.6%; Score 449.6; DB 1; Length 292504;
Best Local Similarity 99.1%; Pred. No. 1.5e-114;
Matches 452; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210815 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 210874

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210875 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 210934

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     210935 TCTTCATTTAATCAGGCAGCCATAATTGGTCAGGCTGGGACTAATAATAGTGCTCAGTTA 210994

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     210995 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAATCGGGCA 211054

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211055 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 211114

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211115 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 211174

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     211175 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAGTTGTAGTGCAGAGA 211234

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||||||||| ||||||||||||||||||||
Db     211235 CAGTCGCAAATGGCAATTCGCGTGACACAACGTTAA 211270
```

```
RESULT 10
CSP515700
LOCUS       CSP515700               2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN
```

Query Match 78.2%; Score 356.8; DB 1; Length 2889;
Best Local Similarity 86.4%; Pred. No. 1.3e-88;
Matches 394; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||| ||
Db       1623 ATGAAAAACAAATTGTTATTTATGATGTTAACAGTACTGGGTGCGCCTGGAATTGCTTCC 1682

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||| || ||||||||||||||  |||| |||||||||||| |||||||||||||| ||||||
Db       1683 GCAACAAGTTATGATTTAGCGCATTCCGAATATAACTTTGCGGTAAATGAATTAAGTAAG 1742

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||||||||||| ||||||||||  || || ||||||||||  || |
Db       1743 TCTTCATTTAATCAGGCAGCCATTATTGGTCAAGTGGGCACGAATAATAGTGCAAAGATG 1802

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
```

```
Db       1803 CGCCAGGAGGGCTCTAAACTTTTGTCGGTAGTTTCACAAGAAGGTGGAAGTAACCGGGCC 1862

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Db       1863 AAAGTTGACCAATCAGGAGCTTATAATTTTGCGTATATTGCTCAGTCGGGTCATTCCAAT 1922

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Db       1923 GATGCCAGTATATCGCAAAGTAATTACGGTAATACTGCGATGATTATCCAGAAAGGTTCT 1982

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Db       1983 GGAAATAAAGCAAATATTACTCAGTATGGTACGCAAAAAACAGCAGTTGTAGTGCAGAGA 2042

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456

Db       2043 CAGTCGCAAATGGCAATTCGCGTTATTCAACGCTAA 2078


RESULT 11
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match           73.0%;  Score 332.8;  DB 1;  Length 2067;
  Best Local Similarity  83.1%;  Pred. No. 6.4e-82;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Db        696 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 755

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Db        756 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 815

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Db        816 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 875

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Db        876 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 935

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Db        936 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 995

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
```

```
                ||||||||||| |||||| | ||||| ||||||| |||     ||||||||||||||||||
Db        996 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 1055

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
                || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       1056 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 1115

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
                |||||||| |||||||||||||| || ||||| |||
Db       1116 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 1151


RESULT 12
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
     gene            3344. .3376
                     /gene="csgBA"
     -35_signal      3344. .3349
                     /gene="csgBA"
     -10_signal      3371. .3376
                     /gene="csgBA"
     gene            3469. .3924
                     /gene="csgB"
     CDS             3469. .3924
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="CAA05316.1"
                     /db_xref="GI:2739238"
                     /db_xref="GOA:P55226"
                     /db_xref="SWISS-PROT:P55226"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     3469. .3531
```

```
                     /gene="csgB"
     mat_peptide     3532. .3921
                     /gene="csgB"
                     /product="CsgB protein"
                     /function="nucleator"
     gene            3966. .4421
                     /gene="csgA"
     CDS             3966. .4421
                     /gene="csgA"
                     /function="fiber subunit"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="CAA05317.1"
                     /db_xref="GI:2739239"
                     /db_xref="GOA:P55225"
                     /db_xref="SWISS-PROT:P55225"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     3966. .4025
                     /gene="csgA"
     mat_peptide     4026. .4418
                     /gene="csgA"
                     /product="CsgA protein"
                     /function="fiber subunit"
     gene            4483. .4809
                     /gene="csgC"
     CDS             4483. .4809
                     /gene="csgC"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgC protein"
                     /protein_id="CAA05318.1"
                     /db_xref="GI:2739240"
                     /db_xref="SWISS-PROT:P55227"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
     sig_peptide     4483. .4533
                     /gene="csgC"
     mat_peptide     4534. .4806
                     /gene="csgC"
                     /product="CsgC"
     gene            4940. .5103
                     /gene="ORF103"
     CDS             4940. .>5103
                     /gene="ORF103"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA05319.1"
                     /db_xref="GI:2739241"
                     /db_xref="SWISS-PROT:O54295"
                     /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                     CDISTHAENI"
ORIGIN

  Query Match             73.0%;  Score 332.8;  DB 1;  Length 5103;
  Best Local Similarity   83.1%;  Pred. No. 6.3e-82;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db       3469 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 3528

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  ||||||| | |||  |||||| ||||| || ||||||||||||| || |||
Db       3529 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 3588

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||| ||||| |||||||||| || ||  |||||||||     ||
Db       3589 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 3648

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db       3649 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 3708

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db       3709 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 3768

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||||||||| |||||| | ||||| ||||||| ||||   ||||||||||||||||||
Db       3769 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 3828

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       3829 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 3888

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||| |||||||||||||| || ||||| |||
Db       3889 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 3924


RESULT 13
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
```

```
        as compressions and repeats; all regions were covered by sequence
        from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .22411
                     /organism="Salmonella typhimurium LT2"
                     /mol_type="genomic DNA"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            434. .1308
                     /gene="phoH"
                     /note="synonym: STM1126"
     RBS             434. .439
                     /gene="phoH"
                     /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS             454. .1308
                     /gene="phoH"
                     /note="similar to E. coli PhoB-dependent, ATP-binding pho
                     regulon component; may be helicase; induced by P
                     starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                     aa), 92% identity in aa 71 - 354"
                     /codon_start=1
                     /transl_table=11
                     /product="PhoB-dependent ATP-binding pho regulon
                     component"
                     /protein_id="AAL20057.1"
                     /db_xref="GI:16419642"
                     /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                     RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                     GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                     YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                     LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                     QRTLHAYS"
     gene            complement(1414. .2303)
                     /gene="STM1127"
     CDS             complement(1414. .2295)
                     /gene="STM1127"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                     identity in aa 15 - 264"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /protein_id="AAL20058.1"
                     /db_xref="GI:16419643"
                     /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                     SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                     SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                     MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                     TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                     QARESKLRTMNALDMTK"
     RBS             complement(2298. .2303)
                     /gene="STM1127"
                     /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene            complement(2580. .4088)
                     /gene="STM1128"
     CDS             complement(2580. .4076)
                     /gene="STM1128"
                     /note="similar to E. coli putative cotransporter
                     (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                     identity in aa 7 - 478"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sodium/glucose cotransporter"
                     /protein_id="AAL20059.1"
                     /db_xref="GI:16419644"
                     /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                     VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                     LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                     LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                     DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                     NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                     AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                     VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                     RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS             complement(4083. .4088)
                     /gene="STM1128"
                     /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene            complement(4413. .5099)
                     /gene="STM1129"
     CDS             complement(4413. .5093)
                     /gene="STM1129"
                     /note="similar to E. coli putative enzyme (AAC76255.1);
                     Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                     226"
                     /codon_start=1
                     /transl_table=11
                     /product="putative inner membrane protein"
                     /protein_id="AAL20060.1"
                     /db_xref="GI:16419645"
                     /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                     AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                     AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                     ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                     WFSHAVKR"
     RBS             complement(5094. .5099)
                     /gene="STM1129"
                     /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene            5599. .6759
                     /gene="STM1130"
     CDS             5599. .6759
                     /gene="STM1130"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                     identity in aa 40 - 403"
                     /codon_start=1
                     /transl_table=11
                     /product="putative inner membrane protein"
                     /protein_id="AAL20061.1"
                     /db_xref="GI:16419646"
                     /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                     AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                     GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                     EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                     PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                     QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                     WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene            6793. .7497
                     /gene="STM1131"
     RBS             6793. .6798
                     /gene="STM1131"
                     /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
```

(AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57% identity in aa 25 - 418"
/codon_start=1
/transl_table=11
/product="putative sugar transport protein"

Query Match 73.0%; Score 332.8; DB 1; Length 22411;
Best Local Similarity 83.1%; Pred. No. 6.2e-82;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db      17272 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 17331

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  ||||||| | |||  |||||| ||||| || ||||||||||||||| || |||
Db      17332 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 17391

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||| ||||| |||||||||  || ||  ||||||||||     ||
Db      17392 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 17451

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| ||||||||| ||| | ||| || | ||||||||  | |  || |||||
Db      17452 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 17511

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db      17512 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 17571

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| ||||   ||||||||||||||||||
Db      17572 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 17631

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db      17632 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 17691

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db      17692 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 17727
```

RESULT 14
AL627269
LOCUS AL627269 254050 bp DNA linear BCT 04-JUL-2003
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome; segment 5/20.
ACCESSION AL627269 AL513382
VERSION AL627269.1 GI:16502231
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi
ORGANISM Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 254050)
AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T., Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N., Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P., Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.
TITLE Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18
JOURNAL Nature 413 (6858), 848-852 (2001)
MEDLINE 21534947
PUBMED 11677608
REFERENCE 2 (bases 1 to 254050)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).

FEATURES Location/Qualifiers

source 1. .254050
/organism="Salmonella enterica subsp. enterica serovar Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"

gene 181. .2793
/gene="STY1078"
/note="synonym: pepN"

CDS 181. .2793
/gene="STY1078"
/note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
/codon_start=1
/transl_table=11
/product="aminopeptidase N"
/protein_id="CAD08183.1"
/db_xref="GI:16502232"
/db_xref="GOA:Q8Z7T0"
/db_xref="SPTREMBL:Q8Z7T0"
/translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
LSGDLYEKITKALA"

misc_feature 532. .1332
/gene="STY1078"
/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase family M1, score 245.10, E-value 1.9e-84"

misc_feature 1060. .1089
/gene="STY1078"
/note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature"

gene 3001. .4011
/gene="STY1079"
/note="synonym: pyrD"

CDS 3001. .4011
/gene="STY1079"
/note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q8Z7S9"
/db_xref="SWISS-PROT:Q8Z7S9"
/translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
HI"

misc_feature 3136. .3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHOdehase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
misc_feature 3238. .3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
misc_feature 3874. .3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
gene 4174. .4719
/gene="STY1080"
CDS 4174. .4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepency in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q8Z7S8"
/translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK IMNDRLKPQVHCHSFSLEQAV"
gene complement(4716. .5825)
/gene="STY1081"
CDS complement(4716. .5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Fasta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q8Z7S7"
/db_xref="SPTREMBL:Q8Z7S7"
/translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
misc_feature complement(4737. .4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
gene 5924. .8032
/gene="STY1082"
CDS 5924. .8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q8Z7S6"
/db_xref="SPTREMBL:Q8Z7S6"
/translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATPAVHFSGLNDTI RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT LSPDFARNRQIHNCWLIRAA"
misc_feature 5927. .7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, score 781.30, E-value 3.7e-231"
misc_feature 6506. .6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
gene 8045. .9952
/gene="STY1083"
CDS 8045. .9952
/gene="STY1083"
/note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa overlap

Query Match 72.6%; Score 331.2; DB 1; Length 254050;
Best Local Similarity 82.9%; Pred. No. 1.7e-81;
Matches 378; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db      88409 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 88468

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  ||||||| | |||  |||||| ||||| || |||||||||||||| || |||
Db      88469 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 88528

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||||| ||||| |||||||||| || ||  |||||||||     ||
Db      88529 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 88588

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||     |  || |||||
Db      88589 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGAAAATAATCGGGCG 88648

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | |||||| |||| ||||||||| |||| |||||||| ||  |||||| |||||||
Db      88649 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 88708

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| ||||   ||||||||||||||||||
Db      88709 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 88768

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db      88769 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 88828

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db      88829 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 88864
```

RESULT 15
AE016840/c
LOCUS AE016840 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of 16 of the complete genome.
ACCESSION AE016840 AE014613
VERSION AE016840.1 GI:29137797
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi Ty2

```
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
                     RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene            complement(5597. .6295)
                     /gene="flgH"
                     /locus_tag="t1740"
     CDS             complement(5597. .6295)
                     /gene="flgH"
                     /locus_tag="t1740"
                     /note="corresponds to STY1219 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar L-ring protein precursor"
                     /protein_id="AAO69364.1"
                     /db_xref="GI:29137803"
                     /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                     PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                     SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                     LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                     MGWLQRFFLNLSPM"
     gene            complement(6350. .7132)
                     /gene="flgG"
                     /locus_tag="t1741"
     CDS             complement(6350. .7132)
```

```
                     /gene="flgG"
                     /locus_tag="t1741"
                     /note="distal rod protein; corresponds to STY1218 from
                     Accession AL513382: Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar basal-body rod protein FlgG"
                     /protein_id="AAO69365.1"
                     /db_xref="GI:29137804"
                     /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                     LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                     FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                     QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                     VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene            complement(7146. .7901)
                     /gene="flgF"
                     /locus_tag="t1742"
     CDS             complement(7146. .7901)
                     /gene="flgF"
                     /locus_tag="t1742"
                     /note="corresponds to STY1217 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="putative flagellar basal-body rod protein FlgF"
                     /protein_id="AAO69366.1"
                     /db_xref="GI:29137805"
                     /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                     PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                     GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                     LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                     DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene            complement(7922. .9133)
                     /gene="flgE"
                     /locus_tag="t1743"
     CDS             complement(7922. .9133)
                     /gene="flgE"
                     /locus_tag="t1743"
                     /note="corresponds to STY1216 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook protein FlgE"
                     /protein_id="AAO69367.1"
                     /db_xref="GI:29137806"
                     /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                     GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL
```

Query Match 72.6%; Score 331.2; DB 1; Length 301983;
Best Local Similarity 82.9%; Pred. No. 1.7e-81;
Matches 378; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db      37807 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 37748

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  ||||||| | |||  |||||| ||||| || ||||||||||||| || |||
Db      37747 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 37688

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||| ||||| |||||||||| || ||  |||||||||     ||
Db      37687 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 37628

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||    |  || |||||
Db      37627 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGAAAATAATCGGGCG 37568

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db      37567 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 37508

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||||||||| |||||| | ||||| ||||||| |||| |||||||||||||||||
Db      37507 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 37448

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db      37447 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 37388

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db      37387 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 37352
```

Search completed: March 18, 2004, 15:02:36
Job time : 2314.32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64620	Aac64620 Escherich
	2	332.8	73.0	456	3	AAC64618	Aac64618 Salmonell
	3	100	21.9	100	7	ACD68810	Acd68810 E. coli K
	4	100	21.9	100	7	ACD68812	Acd68812 E. coli K
	5	100	21.9	100	7	ACD68811	Acd68811 E. coli K
	6	45.8	10.0	779	1	AAN40272	Aan40272 Sequence
	7	38.2	8.4	4860	8	ADA31934	Ada31934 DNA encod
c	8	37.8	8.3	2000	7	ADA71938	Ada71938 Rice gene
	9	37.4	8.2	456	3	AAC64623	Aac64623 AgfA::PT3
	10	36.6	8.0	456	3	AAC64629	Aac64629 AgfA::PT3
c	11	35.2	7.7	4890	2	AAX34648	Aax34648 Promoter
	12	35	7.7	646	2	AAQ62647	Aaq62647 Fibronect
	13	34.6	7.6	361	2	AAQ73066	Aaq73066 AgfA sequ
	14	34.6	7.6	361	2	AAT74141	Aat74141 Salmonell
	15	34.6	7.6	456	2	AAQ87467	Aaq87467 AgfA sequ
	16	34.6	7.6	456	2	AAT74142	Aat74142 Salmonell
	17	34.6	7.6	456	3	AAC64622	Aac64622 AgfA::PT3
	18	34.6	7.6	456	3	AAC64617	Aac64617 Salmonell
	19	34.6	7.6	456	3	AAC64625	Aac64625 AgfA::PT3
	20	34.6	7.6	456	3	AAC64624	Aac64624 AgfA::PT3
c	21	34.4	7.5	4455	7	ABT17054	Abt17054 Human MP2
c	22	34.4	7.5	5184	9	ADE25651	Ade25651 Human cDN
c	23	34.4	7.5	6325	4	AAK80005	Aak80005 Human imm
c	24	34.4	7.5	9555	7	ABX34542	Abx34542 Human mdd
	25	34.2	7.5	1083	5	AAS76745	Aas76745 DNA encod
	26	34	7.5	12426	4	AAS46482	Aas46482 Tumour su
	27	33.8	7.4	1149	3	AAA65947	Aaa65947 E. coli p
	28	33.8	7.4	1149	7	ACA18571	Aca18571 Prokaryot
	29	33.8	7.4	2655	6	ABN70427	Abn70427 Streptoco
	30	33.8	7.4	2655	6	ABN67963	Abn67963 Streptoco
c	31	33.8	7.4	6108	4	ABL08478	Abl08478 Drosophil
c	32	33.8	7.4	6808	4	ABL08488	Abl08488 Drosophil
	33	33.8	7.4	110000	6	ABN71527_06	Continuation (7 of
	34	33.6	7.4	606	3	AAC43625	Aac43625 Arabidops
	35	33.6	7.4	1299	7	ACA48239	Aca48239 Prokaryot
	36	33.2	7.3	1371	7	ACA29389	Aca29389 Prokaryot
	37	33.2	7.3	4590	5	AAH24065	Aah24065 Yeast AOD
	38	33	7.2	1269	7	ACA48121	Aca48121 Prokaryot
	39	33	7.2	1837	7	ACA47731	Aca47731 Prokaryot
	40	33	7.2	110000	6	ABQ69245_18	Continuation (19 o
c	41	33	7.2	319630	6	ABQ67194	Abq67194 Listeria
	42	32.6	7.1	351	7	ABX54413	Abx54413 Bovine ES
	43	32.6	7.1	967	5	AAS74657	Aas74657 DNA encod
c	44	32.6	7.1	1476	6	AAS97253	Aas97253 Neisseria
c	45	32.6	7.1	1476	6	AAS97288	Aas97288 Neisseria

ALIGNMENTS

RESULT 1
AAC64620
ID AAC64620 standard; DNA; 456 BP.
XX
AC AAC64620;
XX
DT 26-FEB-2001 (first entry)
XX
DE Escherichia coli CsgB DNA sequence SEQ ID NO:4.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Escherichia coli.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36344.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 146 A; 78 C; 113 G; 119 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 3.3e-130;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
```

RESULT 2
AAC64618
ID AAC64618 standard; DNA; 456 BP.
XX
AC AAC64618;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis AgfB DNA sequence SEQ ID NO:2.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36342.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 149 A; 87 C; 109 G; 111 T; 0 U; 0 Other;

Query Match 73.0%; Score 332.8; DB 3; Length 456;
Best Local Similarity 83.1%; Pred. No. 3.1e-92;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  |||||| | |||  |||||||||||| || |||||||||||||| || |||
Db         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTGCGGTAAATGAATTAAGCAAG 120

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||| ||||| |||||||||  || ||  ||||||||||      ||
Db        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| |||||||| ||| | ||| || | |||||||||  | |  || |||||
Db        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
```

```
              ||||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456


RESULT 3
ACD68810
ID   ACD68810 standard; DNA; 100 BP.
XX
AC   ACD68810;
XX
DT   18-SEP-2003  (first entry)
XX
DE   E. coli K12 MG1655 biochip probe SEQ ID 80.
XX
KW   Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS   Escherichia coli.
XX
PN   EP1260592-A1.
XX
PD   27-NOV-2002.
XX
PF   17-MAY-2001; 2001EP-00112179.
XX
PR   17-MAY-2001; 2001EP-00112179.
XX
PA   (MWGB-) MWG-BIOTECH AG.
XX
PI   Donner H,  Drescher B,  Huber A,  Weber J;
XX
DR   WPI; 2003-241155/24.
XX
PT   Biochip containing probes complementary with open reading frames in
PT   Escherichia coli K12, useful for detecting gene expression and expression
PT   patterns.
XX
PS   Claim 3; Page 23; 2004pp; German.
XX
CC   This invention describes a novel biochip comprising probe spots, each
CC   containing many identical probes. The probes are nucleotide sequences of
CC   30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC   least one includes a segment of at least 20 bases identical with, or
CC   complementary to, a segment of an open reading frame (orf) of Escherichia
CC   coli K12. The biochip is used for specific detection of gene expression
CC   in K12 and for determining the gene expression pattern, e.g. for
CC   diagnostic determination of which E. coli strains are present in the gut,
CC   and to determine the effects of e.g. growth media on gene expression. The
CC   biochip provides as comprehensive as possible detection of the K12
CC   genome, with simultaneous analysis of many different genes with a single
CC   device, and comparison of gene expression between K12 and its mutants or
CC   other E. coli strains in a single experiment. Apart from qualitative and
CC   quantitative information about gene expression, it also allows
CC   measurements of population densities for the various strains. The use of
CC   synthetic oligonucleotides for preparation of probes allows free
CC   variation in probe length and ensures high purity (and thus selectivity,
CC   reactivity and reproducibility); also synthetic probes are generally
CC   shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC   ACD81540 represent oligonucleotide probes used with the biochip described
CC   in the invention
XX
SQ   Sequence 100 BP; 27 A; 19 C; 26 G; 28 T; 0 U; 0 Other;

  Query Match             21.9%;  Score 100;  DB 7;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-21;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        266 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 60

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 365
              |||||||||||||||||||||||||||||||||||||||
Db         61 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 100


RESULT 4
ACD68812
ID   ACD68812 standard; DNA; 100 BP.
XX
AC   ACD68812;
XX
DT   18-SEP-2003  (first entry)
XX
DE   E. coli K12 MG1655 biochip probe SEQ ID 82.
XX
KW   Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS   Escherichia coli.
XX
PN   EP1260592-A1.
XX
PD   27-NOV-2002.
XX
PF   17-MAY-2001; 2001EP-00112179.
XX
PR   17-MAY-2001; 2001EP-00112179.
XX
PA   (MWGB-) MWG-BIOTECH AG.
XX
PI   Donner H,  Drescher B,  Huber A,  Weber J;
XX
DR   WPI; 2003-241155/24.
XX
PT   Biochip containing probes complementary with open reading frames in
PT   Escherichia coli K12, useful for detecting gene expression and expression
PT   patterns.
XX
PS   Claim 3; Page 23; 2004pp; German.
XX
CC   This invention describes a novel biochip comprising probe spots, each
CC   containing many identical probes. The probes are nucleotide sequences of
CC   30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC   least one includes a segment of at least 20 bases identical with, or
CC   complementary to, a segment of an open reading frame (orf) of Escherichia
CC   coli K12. The biochip is used for specific detection of gene expression
CC   in K12 and for determining the gene expression pattern, e.g. for
CC   diagnostic determination of which E. coli strains are present in the gut,
CC   and to determine the effects of e.g. growth media on gene expression. The
CC   biochip provides as comprehensive as possible detection of the K12
CC   genome, with simultaneous analysis of many different genes with a single
CC   device, and comparison of gene expression between K12 and its mutants or
CC   other E. coli strains in a single experiment. Apart from qualitative and
CC   quantitative information about gene expression, it also allows
CC   measurements of population densities for the various strains. The use of
CC   synthetic oligonucleotides for preparation of probes allows free
CC   variation in probe length and ensures high purity (and thus selectivity,
CC   reactivity and reproducibility); also synthetic probes are generally
CC   shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC   ACD81540 represent oligonucleotide probes used with the biochip described
CC   in the invention
XX
SQ   Sequence 100 BP; 28 A; 18 C; 27 G; 27 T; 0 U; 0 Other;

  Query Match             21.9%;  Score 100;  DB 7;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-21;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        123 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 182
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 60

Qy        183 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 222
                 |||||||||||||||||||||||||||||||||||||||
Db         61 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 100
```

RESULT 5
ACD68811
ID ACD68811 standard; DNA; 100 BP.
XX
AC ACD68811;
XX
DT 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 81.
XX
KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EP1260592-A1.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MWGB-) MWG-BIOTECH AG.
XX
PI Donner H, Drescher B, Huber A, Weber J;
XX
DR WPI; 2003-241155/24.
XX
PT Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX
PS Claim 3; Page 23; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SQ Sequence 100 BP; 31 A; 20 C; 29 G; 20 T; 0 U; 0 Other;

  Query Match   21.9%; Score 100; DB 7; Length 100;
  Best Local Similarity   100.0%; Pred. No. 9.5e-21;
  Matches 100; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy        208 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 267
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 60

Qy        268 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 307
                 |||||||||||||||||||||||||||||||||||||||
Db         61 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 100
```

RESULT 6
AAN40272
ID AAN40272 standard; cDNA; 779 BP.
XX
AC AAN40272;
XX
DT 09-JAN-1992 (first entry)
XX
DE Sequence of a clone numbered as Ag16 which corresponds to the S antigen
DE of isolate FC27.
XX
KW Malaria; vaccine; antigen; epitope; immune response; ss.
XX
OS Plasmodium falciparum.
XX
FH Key   Location/Qualifiers
FT CDS   1. .777
FT   /*tag= d
FT misc_feature   1. .8
FT   /*tag= a
FT   /label= linker
FT repeat_region   10. .768
FT   /*tag= c
FT   /rpt_unit= 10. .42
FT   /note= "final base of certain codons may differ but AA SQ
FT   encoded stays the same"
FT misc_feature   772. .779
FT   /*tag= b
FT   /label= linker
XX
PN WO8402917-A.
XX
PD 02-AUG-1984.
XX
PF 27-JAN-1984; 84WO-AU000016.
XX
PR 28-JAN-1983; 83AU-00007843.
PR 10-JUN-1983; 83AU-00009788.
PR 26-JAN-1984; 84AU-00023842.
XX
PA (HALL-) HALL INST MED RES.
PA (KEMP/) KEMP D J.
PA (AUME-) AUST INST MED RES.
PA (SARA-) SARAMANE PTY LTD.
XX
PI Kemp DJ, Anders RF, Coppel RL, Brown G, Saint R, Cowman AF;
PI Mitchell GF;
XX
DR WPI; 1984-201410/32.
DR P-PSDB; AAP40376.
XX
PT Expression of plasmodium falciparum poly:peptide(s) from CDNA - for use
PT in immunisation against malaria infection.
XX
PS Claim 5; Fig 1; 62pp; English.
XX
CC The inventors claim a DNA SQ which comprises a nucleotide sequence
CC corresp. to all or part of Plasmodium falciparum RNA. Also claimed is a
CC synthetic peptide or polypeptide displaying the antigenicity of all or
CC part of a P.falciparum antigen. The base sequence of clone Ag16 (see
CC AAN40272) indicates that the antigen coded for by this partial
CC polynucleotide sequence has a homologous repeat structure of 11 AAs
CC tandemly repeated 23 times
XX
SQ Sequence 779 BP; 294 A; 125 C; 217 G; 143 T; 0 U; 0 Other;

  Query Match   10.0%; Score 45.8; DB 1; Length 779;

Best Local Similarity 47.7%; Pred. No. 0.0011;
Matches 134; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```
Qy        126 ATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCA 185
              ||| |  |  | || | |  | ||||| |||      |  ||  |||  ||  | | ||
Db        198 ATTAGAAGATCCTGCAAAAGCTAGTCAAGGTGGATTAGAAGATCCTGCAAAGGCAAGTCA 257

Qy        186 GGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGAT 245
              |||||  | | ||  |  ||    | |   ||||  ||         | |  |||||
Db        258 AGGAGGATTAGAAGATCCTGCAAAAGCTAGTCAAGGTGGATTAGAAGATCCTGCAAAAGC 317

Qy        246 TGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGC 305
              |   ||  ||||   |  | | |||| |   |  ||| |  ||    |   |   |||
Db        318 TAGTCAAGGAGGATTAGAAGATCCTGCAAAAGCTAGTCAAGGTGGATTAGAAGATCCTGC 377

Qy        306 CAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 365
              |   |  |||||||  |  |   || | || | |  ||  ||    || |  |   |
Db        378 AAAAGCTAGTCAAGGTGGATTAGAAGATCCCGCAAAGGCTAGTCAAGGTGGATTAGAAGA 437

Qy        366 TAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAA 406
              |  |||||  ||  ||  |||     || |  | ||||
Db        438 TCCTGCAAAAGCTAGTCAAGGTGGATTAGAAGATCCCGCAA 478
```

RESULT 7
ADA31934
ID ADA31934 standard; DNA; 4860 BP.
XX
AC ADA31934;
XX
DT 20-NOV-2003 (first entry)
XX
DE DNA encoding Acinetobacter baumannii protein #3221.
XX
KW ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KW vaccine; plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR P-PSDB; ADA36060.
XX
PT New Acinetobacter baumanii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 3221; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents DNA encoding an A. baumannii
CC protein.
XX
SQ Sequence 4860 BP; 1639 A; 752 C; 998 G; 1471 T; 0 U; 0 Other;

Query Match 8.4%; Score 38.2; DB 8; Length 4860;
Best Local Similarity 45.9%; Pred. No. 0.51;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy        106 AATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAAT 165
              |||    |  |||| |  | |  || | | |||||    | |  |||||| |   || |
Db       3445 AATACTATTGGTAAATTATTAGGTACTGATGCAGCAGATAATCTTCAAGCAGATGCTGAG 3504

Qy        166 AATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGT 225
              | |   ||  |   |  |||||||    |     |   | || ||  ||  | | |
Db       3505 ATTTCAACTATTTATGGTTTGGGAGGCAATGATACGATTCAAGGTGGTGTACAGAATGAT 3564

Qy        226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
              |     | | |    ||| | ||| || || |||    | | |  | || |
Db       3565 TATCTGTATGGTGATGATGGAGACGATACATTAGTTTCTAATACAGGCTCTGATTATTTG 3624

Qy        286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
              ||  ||||    ||| ||  |    | ||| | | ||  |||      | | |
Db       3625 TATGGTGGTGCAGGTAATGACACATTGATTTATGGCGGTAATTCCAATGTGTATACGGCT 3684

Qy        346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATG 388
              | ||      |||| ||| |  |  |||||| | |  | |
Db       3685 TTACAAGGTCAAGCTGGGAATGACACTTATATTATAGATAAAG 3727
```

RESULT 8
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX

```
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

  Query Match              8.3%;  Score 37.8;  DB 7;  Length 2000;
  Best Local Similarity   13.1%;  Pred. No. 0.47;
  Matches   36;  Conservative  121;  Mismatches  118;  Indels    0;  Gaps    0;

Qy         21 TATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATGATTTAGC 80
Db        950 WAAGRWRWRMAWCWYCCMWKWWKMTSCMWWKYWRTWSCWYTMWWGAMRYAYYAMRRRRWT 891

Qy         81 TAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGC 140
Db        890 YKWSWRRMYWTMTKWAWTWMTCMCMAKWYMATGWATWMWWRYTMYTYCYAMTCAKCKYKM 831

Qy        141 CATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACT 200
Db        830 AMTKWWTTWACAWRATSWRWRAMAGMRWKRYKMKRAYWWRWWRCWKAGWARWMKSRYRWK 771

Qy        201 TTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGA 260
Db        770 WKKYATRYYWKMWAMTWWWSWRRWKSYRMWSGMGRMRWSAWRYCSRMKCAKTKYASSARW 711

Qy        261 TTATAACCTTGCATATATTGATCAGGCGGGCAGTG 295
Db        710 TKRAKRSYRYRRRWYWKRKGWTYRYRYWRSCRMTR 676


RESULT 9
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match              8.2%;  Score 37.4;  DB 3;  Length 456;
  Best Local Similarity   51.5%;  Pred. No. 0.35;
  Matches   86;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
Db        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
Db        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327

Qy        397 AAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGGCTATTCGCGT 443
Db        328 AACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGT 374


RESULT 10
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
```

XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 8.0%; Score 36.6; DB 3; Length 456;
Best Local Similarity 54.0%; Pred. No. 0.61;
Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy        275 ATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATA 334
              | || |  || || ||   ||  ||||  |||  | |    || |||||  ||  || ||
Db        206 AAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTA 265

Qy        335 CTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTC 394
              ||       | |  ||||| ||||   | ||||| || |  ||    ||||      ||
Db        266 CTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325

Qy        395 AAAAAACGGCAATTGTAGT 413
              ||||  | |  |||   ||
Db        326 AAAACTCCGATATTACTGT 344
```

RESULT 11
AAX34648/c
ID AAX34648 standard; DNA; 4890 BP.
XX
AC AAX34648;
XX
DT 17-OCT-2003 (revised)
DT 05-JUL-1999 (first entry)
XX
DE Promoter containing sequence of SBE I.
XX
KW Starch biosynthetic pathway; cereal plant; enzyme; SBE; SSS; DBE; GBSS;
KW starch branching enzyme; starch soluble synthase; debranching enzyme;
KW endosperm; wheat; barley; granule-bound synthase; glutenin; starch;
KW grain softness protein I; bacterial isoamylase; glycogen synthase;
KW wSBE I-D4 gene; promoter; ss.
XX
OS Aegilops tauschii.
XX
PN WO9914314-A1.
XX
PD 25-MAR-1999.
XX
PF 11-SEP-1998; 98WO-AU000743.
XX
PR 12-SEP-1997; 97AU-00009108.
PR 20-MAR-1998; 98AU-00002509.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA (AUSU) UNIV AUSTRALIAN NAT.
PA (GOOD-) GOODMAN FIELDER LTD.
PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.
XX
PI Li Z, Morell M, Rahman S;
XX
DR WPI; 1999-229525/19.
XX
PT New isolated cereal plant enzyme genes used for, e.g. expression of
PT antisense sequences of granule bound synthase.
XX
PS Claim 17; Page 63-65; 171pp; English.
XX
CC The invention relates to a novel enzyme of starch biosynthetic pathway in
CC a cereal plant, where the enzyme is selected from starch branching enzyme
CC (SBE) I , SBE II, starch soluble synthase (SSS) I, and debranching enzyme
CC (DBE), with the proviso that the enzyme is not SSS I of rice, or SBE I of
CC rice or maize. The methods and products can be used for targeting
CC expression specifically to the endosperm of the seeds of cereal plants
CC such as wheat or barley. They can be used for the expression of e.g.
CC antisense sequences of granule-bound synthase (GBSS), SBE II, low mol.
CC wt. glutenin, grain softness protein I, bacterial isoamylase, bacterial
CC glycogen synthase, and wheat high mol. wt. glutenin Bx17. They can be
CC used for modifying the characteristics of starch produced by a plant. The
CC present sequence represents the wheat SBE I promoter containing sequence.
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 4890 BP; 1403 A; 1067 C; 1035 G; 1385 T; 0 U; 0 Other;

Query Match 7.7%; Score 35.2; DB 2; Length 4890;
Best Local Similarity 55.8%; Pred. No. 4.3;
Matches 87; Conservative 0; Mismatches 68; Indels 1; Gaps 1;

```
Qy        258 AGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCA 317
              |||  | ||| || |||| | ||| ||  |   |||||||||  ||||||  |  || |||
Db       1097 AGAGAAGAACTTTCCATAAAATGAACAAACAAACAGTGCCAAGTATGCCACTTGTTGGCA 1038

Qy        318 AGGTGCTTATGGTAATACTGCGA-TGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATA 376
              |       ||   || | ||  | | |       |  ||||||||  |  |  ||||   |
Db       1037 ACTCATACATTTGAAAATTGTTAGTTAAATGAAACTAAGGTTCTATTCTTTTAGCATGCA 978

Qy        377 TTACACAGTATGGTACTCAAAAAACGGCAATTGTAG 412
              || ||  | ||        | |  ||  ||||||||
Db        977 TTGCATTGCATTACTAGTTAGACCCGTGAATTGTAG 942
```

RESULT 12
AAQ62647
ID AAQ62647 standard; DNA; 646 BP.
XX
AC AAQ62647;
XX
DT 25-MAR-2003 (revised)
DT 18-OCT-1994 (first entry)
XX
DE Fibronectin binding protein fnbA gene.
XX
KW Fibronectin binding protein; FBP; curlin; curli pili; pilus;
KW immunization; vaccine; fnbA gene; ss.
XX
OS Escherichia coli.
XX
PN WO9410330-A1.
XX
PD 11-MAY-1994.
XX
PF 03-NOV-1993; 93WO-US010547.
XX

```
PR   03-NOV-1992;   92US-00970846.
XX
PA   (REAT/) REA T S.
XX
PI   Normark S,  Olsen A;
XX
DR   WPI; 1994-167484/20.
DR   P-PSDB; AAR52664.
XX
PT   New fibronectin binding protein curlin from E. coli - used for the
PT   treatment of infections caused by E. coli and for immunisation against E.
PT   coli infection.
XX
PS   Disclosure; Fig 2; 33pp; English.
XX
CC   The fnbA gene of E. coli cattle fecal isolate AO12 includes DNA encoding
CC   a new fibronectin binding protein (FBP), curlin, which is produced as
CC   curli pili when expressed in E. coli (pFnb20) DSM 4585. The fnbA gene
CC   sequence from pFnb20 is given in AAQ62647, and includes the sequence
CC   encoding the 17 kDa curlin subunit (amino acids 21-142 of AAR52664).
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 646 BP; 168 A; 146 C; 161 G; 171 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 35;  DB 2;  Length 646;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        416 AGAGACAGTCGCAAATGGCTATTCGCGTGACACAA 450
              ||||||||||||||||||||||||||||||||||
Db          1 AGAGACAGTCGCAAATGGCTATTCGCGTGACACAA 35


RESULT 13
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             7.6%;  Score 34.6;  DB 2;  Length 361;
  Best Local Similarity   53.3%;  Pred. No. 2.3;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              |||   |||   ||   ||  ||||  |||  | |     || |||||  ||  || ||||
Db        145 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 204

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
                  | |  ||||| ||||   | ||||| || |  ||     ||||          || ||
Db        205 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 264

Qy        397 AAAACGGCAATTGTAGT 413
              ||  | |  |||   ||
Db        265 AACTCCGATATTACTGT 281


RESULT 14
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
```

```
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             7.6%;  Score 34.6;  DB 2;  Length 361;
  Best Local Similarity   53.3%;  Pred. No. 2.3;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              |||   |||   ||   ||  ||||  |||  | |    || |||||  ||  || ||||
Db        145 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 204

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
                | |  ||||| ||||   | ||||| || |  ||     ||||       || ||
Db        205 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 264

Qy        397 AAAACGGCAATTGTAGT 413
              ||  | |  |||   ||
Db        265 AACTCCGATATTACTGT 281


RESULT 15
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag=  a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             7.6%;  Score 34.6;  DB 2;  Length 456;
  Best Local Similarity   53.3%;  Pred. No. 2.5;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              |||   |||   ||   ||  ||||  |||  | |    || |||||  ||  || ||||
Db        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
                | |  ||||| ||||   | ||||| || |  ||     ||||       || ||
Db        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327

Qy        397 AAAACGGCAATTGTAGT 413
              ||  | |  |||   ||
Db        328 AACTCCGATATTACTGT 344


Search completed: March 18, 2004, 09:11:52
Job time : 291.405 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	39.6	8.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	38.2	8.4	4860	4	US-09-328-352-3221	Sequence 3221, Ap
	3	34.6	7.6	361	1	US-08-233-788A-56	Sequence 56, Appl
	4	34.6	7.6	456	1	US-08-233-788A-58	Sequence 58, Appl
	5	31.8	7.0	580073	4	US-08-545-528D-1	Sequence 1, Appli
	6	31.8	7.0	640681	4	US-09-790-988-1	Sequence 1, Appli
	7	31.4	6.9	525	4	US-09-107-532A-2325	Sequence 2325, Ap
c	8	31.4	6.9	2785	4	US-08-956-171E-133	Sequence 133, App
	9	31.4	6.9	53332	4	US-09-801-861-3	Sequence 3, Appli
c	10	31.4	6.9	62909	4	US-09-596-002-32	Sequence 32, Appl
	11	31.2	6.8	606	4	US-09-540-236-1518	Sequence 1518, Ap
	12	31.2	6.8	1404	4	US-09-134-001C-398	Sequence 398, App
	13	31	6.8	2139	4	US-09-059-584-50	Sequence 50, Appl
c	14	31	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	15	30.6	6.7	832	4	US-09-621-976-2813	Sequence 2813, Ap
	16	30.6	6.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	17	30.4	6.7	831	4	US-09-328-352-3901	Sequence 3901, Ap
	18	30.4	6.7	1023	4	US-09-540-236-1556	Sequence 1556, Ap
c	19	30.4	6.7	1227	4	US-09-107-532A-2544	Sequence 2544, Ap
c	20	30.4	6.7	1622	4	US-08-956-171E-732	Sequence 732, App
	21	30.4	6.7	1776	4	US-09-134-000C-3324	Sequence 3324, Ap
	22	30.4	6.7	1879	4	US-09-023-655-661	Sequence 661, App
	23	30.4	6.7	2001	3	US-08-714-918-92	Sequence 92, Appl
	24	30.4	6.7	2001	3	US-09-265-315-92	Sequence 92, Appl
	25	30.4	6.7	2001	3	US-09-265-315-92	Sequence 92, Appl
	26	30.4	6.7	2001	3	US-09-266-417-92	Sequence 92, Appl
	27	30.4	6.7	2001	4	US-09-528-709-92	Sequence 92, Appl
	28	30.4	6.7	2001	4	US-09-527-745-92	Sequence 92, Appl
	29	30.4	6.7	99629	4	US-09-596-002-37	Sequence 37, Appl
c	30	30.2	6.6	417	4	US-09-621-976-10111	Sequence 10111, A
	31	30.2	6.6	2364	3	US-09-172-045-1	Sequence 1, Appli
	32	30.2	6.6	2364	4	US-09-342-325C-1	Sequence 1, Appli
	33	30.2	6.6	4215	4	US-09-620-312D-295	Sequence 295, App
	34	30.2	6.6	13508	4	US-08-956-171E-120	Sequence 120, App
	35	30.2	6.6	152331	3	US-09-128-155-16	Sequence 16, Appl
	36	30.2	6.6	176373	3	US-09-128-155-17	Sequence 17, Appl
	37	30	6.6	2136	2	US-08-949-941B-1	Sequence 1, Appli
c	38	29.8	6.5	7091	3	US-08-975-762-46	Sequence 46, Appl
c	39	29.8	6.5	7091	3	US-09-295-028-46	Sequence 46, Appl
c	40	29.8	6.5	7091	4	US-09-106-582-46	Sequence 46, Appl
c	41	29.8	6.5	7091	4	US-09-159-469-46	Sequence 46, Appl
c	42	29.8	6.5	7091	4	US-09-693-542-46	Sequence 46, Appl
c	43	29.8	6.5	13158	2	US-08-687-080-105	Sequence 105, App
c	44	29.4	6.4	703	4	US-09-016-434-120	Sequence 120, App
c	45	29.2	6.4	369	4	US-09-328-352-3808	Sequence 3808, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 8.7%; Score 39.6; DB 1; Length 7218;

```
  Best Local Similarity   10.5%;  Pred. No. 0.0088;
  Matches   39;  Conservative  166;  Mismatches  165;  Indels    0;  Gaps    0;

Qy          64 GCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCT 123
               |||  |  | | ||    ||  | | || ||           ||| | |  |  || |
Db        1523 GCATCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTACCTATCTATGCAA 1464

Qy         124 TCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGG 183
                 |||| ||  ||    | ||||| :: :::: :: :: :: :   ::  : ::
Db        1463 GTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1404

Qy         184 CAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAG 243
                :::::::   ::::     :: ::  :  : : ::::::: :: :: ::  ::: ::::
Db        1403 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1344

Qy         244 ATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGAT 303
               :  ::  ::: ::::::  : ::    : : : :  ::  ::: ::: :: :  :: ::
Db        1343 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1284

Qy         304 GCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGT 363
               :  :: :    : :::: :   : :: :: :  : :: ::  :   :::::::    ::
Db        1283 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1224

Qy         364 AATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAG 423
              :: :::: ::: :  : : :: : :: :   :::::: :: ::  : :: : ::::: ::
Db        1223 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1164

Qy         424 TCGCAAATGG 433
                 : ::: ::
Db        1163 RRRRRRRRRR 1154


RESULT 2
US-09-328-352-3221
; Sequence 3221, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3221
;   LENGTH: 4860
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-3221

  Query Match             8.4%;  Score 38.2;  DB 4;  Length 4860;
  Best Local Similarity   45.9%;  Pred. No. 0.021;
  Matches  130;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy         106 AATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAAT 165
               |||    | |||| |  | |  || | | |||||    | |  |||||| |   || |
Db        3445 AATACTATTGGTAAATTATTAGGTACTGATGCAGCAGATAATCTTCAAGCAGATGCTGAG 3504

Qy         166 AATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGT 225
               | |    ||   |   |  |||||||     |      |    | || ||  ||  | | |
Db        3505 ATTTCAACTATTTATGGTTTGGGAGGCAATGATACGATTCAAGGTGGTGTACAGAATGAT 3564

Qy         226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
                 |    | | |   |  ||| | |||  || || |||    |  | |  | ||  |
Db        3565 TATCTGTATGGTGATGATGGAGACGATACATTAGTTTCTAATACAGGCTCTGATTATTTG 3624

Qy         286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
                ||  ||||    ||| ||    |      | || | | ||   |||      | |  |
Db        3625 TATGGTGGTGCAGGTAATGACACATTGATTTATGGCGGTAATTCCAATGTGTATACGGCT 3684

Qy         346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATG 388
                | ||         |||| ||| |  |  |||||| | |  | |
Db        3685 TTACAAGGTCAAGCTGGGAATGACACTTATATTATAGATAAAG 3727


RESULT 3
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             7.6%;  Score 34.6;  DB 1;  Length 361;
  Best Local Similarity   53.3%;  Pred. No. 0.087;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
               |||   |||   ||   ||  ||||  |||  | |    || |||||  ||  || ||||
Db         145 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 204

Qy         337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
                   | |  |||||  ||||   | |||||  || |  ||    ||||       || ||
Db         205 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 264

Qy         397 AAAACGGCAATTGTAGT 413
               ||  | |  |||   ||
Db         265 AACTCCGATATTACTGT 281


RESULT 4
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
```

```
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-788A-58

  Query Match             7.6%;  Score 34.6;  DB 1;  Length 456;
  Best Local Similarity   53.3%;  Pred. No. 0.098;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              ||| |||  ||  ||  |||| ||| | |   || |||||  ||  || ||||
Db        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
                | |  ||||| ||||   | ||||| || |  ||    ||||      || ||
Db        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327

Qy        397 AAAACGGCAATTGTAGT 413
              ||  | |  |||  ||
Db        328 AACTCCGATATTACTGT 344


RESULT 5
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
;  APPLICANT: Fraser et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB193P1
;  CURRENT APPLICATION NUMBER: US/08/545,528D
;  CURRENT FILING DATE:  1995-10-19
;  PRIOR APPLICATION NUMBER: US 08/488,018
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/473,545
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 580073
;   TYPE: DNA
;   ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

  Query Match             7.0%;  Score 31.8;  DB 4;  Length 580073;
  Best Local Similarity   43.1%;  Pred. No. 30;
  Matches  153;  Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

Qy         26 TGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATGATTTAGCTAATT 85
              | ||||      |||      | |  | | |||    | ||  | |  | |||||  ||
Db     181203 TTTTAAATGGTTTGGCACTAACAGTTAATTCAGTTTTAACAAATGAACAGTTAGCAGATG 181262

Qy         86 CAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAA 145
              ||      |||||| | | |  |      ||  |  ||  | |||  || | |   ||
Db     181263 CATGTGTTAACTTTGGGTTTGACTTTAAGATGGAAACTGAAGTTACCCATGAAAATATTG 181322

Qy        146 TTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGG 205
              | | | |       |  | |  ||||| | | |||  |||   | |  ||    || |
Db     181323 TAGCTAACATCCAGTTTGAAGATAGTGATGATTTATTGCAACCAAGACCACCTATTGTTA 181382

Qy        206 CGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATA 265
              |  | | |  || |  |    | | ||      |      | |||   |   || |  ||
Db     181383 CTATCATGGGTCATGTTGACCATGGTAAAACTTCGCTTTTAGACACAATTAGAAAAACTA 181442

Qy        266 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 325
              |  |  |   ||  ||    |  || | | || |  |     ||  ||  || |
Db     181443 ATGTAACTGCTAAGGAGTTTGGCGGAATTACCCAAAAAATTGGTGCTTATCAGGTGAAAA 181502

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTAC 380
              ||    | | |   | |||||      |  | | ||  |    | | |    ||||
Db     181503 ATCACCAAAATAAAACGATTACTTTTATTGATACTCCTGGGCATGAAGCATTTAC 181557


RESULT 6
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             7.0%;  Score 31.8;  DB 4;  Length 640681;
  Best Local Similarity   46.9%;  Pred. No. 32;
  Matches   99;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCG 61
              | |  ||  ||  || |   || |  ||||||  |       ||   ||  |  |    |
Db      79853 TTAGCAAAGAAACGTCAGAGATAAATTTAACACAAAAAAAAGCATATGAAACAGGTTTCC 79912

Qy         62 CAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGT 121
               | ||||   |     ||  ||    ||  | |  ||    |  |||| |||| | | | |
Db      79913 AAACAGGACTTCTAAAAGGCAAAGAGGAGAACACTTTGTTGAAAAATAAATTAAATGATT 79972
```

```
Qy        122 CTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTAC 181
              |  || |||  | |       ||| | ||  ||  | | |     || |||| ||||
Db      79973 TATTTTTAAATTTTGAAAATTCTCTTGCTAAATTTGAAAATGAACTATGTTCTCAATTAC 80032

Qy        182 GGCAGGGAGGCTCAAAACTTTTGGCGGTTGT 212
              |   ||  | |||| |||   |   |||
Db      80033 TAAAAACAGTTTTAAAAATTTCATCTTATGT 80063


RESULT 7
US-09-107-532A-2325
; Sequence 2325, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;    INFORMATION FOR SEQ ID NO: 2325:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 525 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...525
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2325:
US-09-107-532A-2325

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 525;
  Best Local Similarity   54.9%;  Pred. No. 1.2;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        270 TGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGG 329
              || ||    |||| |||||       |     | ||||| ||||||   |||   |  ||
Db        213 TGAATTATCTGATAAGGCGAAAGAAGAAGCAGTTGCCATTATTTCTTCAGGGATTAGTGA 272

Qy        330 TAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACAC 382
                ||| ||  |||| |    | |    || |||  ||||||  || |||  ||
Db        273 CTATATTGATATGAGTGGTGAAATCAATTATGGCGATAAAGACAAAATTGTAC 325


RESULT 8
US-08-956-171E-133/c
; Sequence 133, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/956,171E
;             FILING DATE: 20-Oct-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;    INFORMATION FOR SEQ ID NO: 133:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2785 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 133:
US-08-956-171E-133

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2785;
  Best Local Similarity   49.7%;  Pred. No. 2.8;
  Matches   80;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        125 CATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGC 184
              || ||||||| ||||  || |||  | || | |    ||  |  ||||| |||| |||
Db       2531 CAATTAATCAAGCAGATATCATTTCTGAACCGGTTGCTCTTGGTAGTGTTCAGGTACCAA 2472

Qy        185 AGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGA 244
              | |  ||   |    || |   | ||  ||   || ||   |   |         | ||| |
Db       2471 ATGATGGTCAACCTATTATTCTGCTTAATGATAAACAAACGATAGGTGGTTATACCAAAA 2412

Qy        245 TTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
              |||       | |  | | | | ||  || |  | |  |||
Db       2411 TTGCAACTGTATGCAAATTTGATCTGCCAAAGTTAGCACAG 2371


RESULT 9
US-09-801-861-3
; Sequence 3, Application US/09801861
; Patent No. 6492154
```

; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001098
; CURRENT APPLICATION NUMBER: US/09/801,861
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 53332
; TYPE: DNA
; ORGANISM: Human
US-09-801-861-3

Query Match 6.9%; Score 31.4; DB 4; Length 53332;
Best Local Similarity 61.7%; Pred. No. 12;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 336 TGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCA 395
Db 8632 TGTTATGAGTATTCAACAAGAATATGTATATAAAGCACTTATCACAGAGTCAGTTTTTCA 8691

Qy 396 AAAAACGGCAATTGTAGTGCA 416
Db 8692 GAGTTCAACAAATGTTGACCA 8712

RESULT 10
US-09-596-002-32/c
; Sequence 32, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 62909
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 32
; PUBLICATION INFORMATION:
US-09-596-002-32

Query Match 6.9%; Score 31.4; DB 4; Length 62909;
Best Local Similarity 49.7%; Pred. No. 14;
Matches 80; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 221 AAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTG 280
Db 8029 AATTGATTAACAACCATATGGAGTATGTTATGAAAAGATTATCTTTAATTGCATTATTGG 7970

Qy 281 ATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGA 340
Db 7969 CGGTGGCAGGTTTTGCTAACGCAGCGACTTATGAGCTTGATCCTTATCACACCAATGCGC 7910

Qy 341 TGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACA 381
Db 7909 GTTTTAGCATTGACCATTTTGGCACCTCAACCAATGTGGCA 7869

RESULT 11
US-09-540-236-1518
; Sequence 1518, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1518
; LENGTH: 606
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-1518

Query Match 6.8%; Score 31.2; DB 4; Length 606;
Best Local Similarity 50.0%; Pred. No. 1.5;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
Db 4 ATTAACAACCATATGGAGTATGTTATGAAAAGATTATCTTTAATTGCATTATTGGCGGTG 63

Qy 286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
Db 64 GCAGGTTTTGCTAACGCAGCGACTTATGAGCTTGATCCTTATCACACCAATGCGCGTTTT 123

Qy 346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACA 381
Db 124 AGCATTGACCATTTTGGCACCTCAACCAATGTGGCA 159

RESULT 12
US-09-134-001C-398
; Sequence 398, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 398
; LENGTH: 1404
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-398

Query Match 6.8%; Score 31.2; DB 4; Length 1404;
Best Local Similarity 52.3%; Pred. No. 2.3;
Matches 69; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 275 ATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATA 334
Db 1079 ATTCTGATAAGAACGCAGATGATGACTCTGATAATAGTTCCTCAAATCCTAACGCTACTT 1138

Qy 335 CTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTC 394
Db 1139 CAACTAATAATAACGATAATGTAGCCAATAATAACTCAAATTATACAAACCAAAATCAAC 1198

Qy 395 AAAAAACGGCAA 406
Db 1199 AAGATAATGCAA 1210

```
RESULT 13
US-09-059-584-50
; Sequence 50, Application US/09059584
; Patent No. 6440701
;  GENERAL INFORMATION:
;     APPLICANT:  Myers, Lisa E
;     APPLICANT:  Schryvers, Anthony B
;     APPLICANT:  Harkness, Robin E
;     APPLICANT:  Loosmore, Sheena M.
;     APPLICANT:  Du, Run-Pan
;     APPLICANT:  Yang, Yan-Ping
;     APPLICANT:  Klein, Michel H
;     TITLE OF INVENTION:  Transferrin Receptor Genes of Moraxella
;     NUMBER OF SEQUENCES:  60
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sim & McBurney
;       STREET:  6th Floor, 330 University Avenue
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/059,584
;       FILING DATE:  14-APR-1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/778,570
;       FILING DATE:  03-JAN-1997
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Stewart, Michael I
;       REGISTRATION NUMBER:  24973
;       REFERENCE/DOCKET NUMBER:  1038-794
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 595-1155
;       TELEFAX:  (416) 595-1163
;   INFORMATION FOR SEQ ID NO:  50:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2139 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-059-584-50

  Query Match             6.8%;  Score 31;  DB 4;  Length 2139;
  Best Local Similarity   59.8%;  Pred. No. 3.3;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy        306 CAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 365
              ||||  |   ||  ||||| || || |||||| |  |||       || |||| | ||
Db       1845 CAGTGATGGTAAATCTGCTTTTGATATTACTGCTACAATTGATGGCAATGGTTTTAGTGG 1904

Qy        366 TAAAGCAAATATTACACAGTATGGTAC 392
              |||||| ||||   || |   ||  ||
Db       1905 TAAAGCCAATACACCAGATATTGAAAC 1931


RESULT 14
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             6.8%;  Score 31;  DB 4;  Length 1664976;
  Best Local Similarity   49.1%;  Pred. No. 83;
  Matches   82;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy        239 CAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCA 298
              || | | | |  ||   |  ||   || | | ||  |  ||| || || || ||  |
Db     122859 CATATAATAAGGAGGATGATGAAACTACCGTGGCTACTTATGACCATGCAGGGAGGACTG 122800

Qy        299 ACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTT 358
              |||  || |  | |  |   |  | ||| | |  |||     |  ||   || |  |
Db     122799 GAGATTACAATTATGCAAATAATATTGATGCTTACGCTGGTTCAAAAATAGCGATAAAAT 122740

Qy        359 CTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCA 405
              |   | |  ||||   |  | |||||||||   |  |  ||| |||
Db     122739 ATACAAGTGGAGCATTAACAAGACAGTATGGGGTTTTAGGAACTGCA 122693


RESULT 15
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 832;
  Best Local Similarity   16.5%;  Pred. No. 2.8;
  Matches   69;  Conservative  151;  Mismatches  195;  Indels    2;  Gaps    1;

Qy         10 AAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGT 69
              ::::::: :::::: :| :::::|::  ::::  |:  |::: ::|:: :      ::
Db         40 WRRKKKKAWWKYWKTWTWWYWRYAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYM 99

Qy         70 TATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTT 129
               :: :|:|   : :  :  :::: :  : :  : |   : :|:  ::   : |: :::|
Db        100 AMWTRTWTGYAYYRSMMYWWRYRCWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKT 159
```

```
Qy        130 AATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGA 189
Db        160 YYWAATRYWWMMCWTKRWRASWWYCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGA 219

Qy        190 GGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGAC 249
Db        220 MSWKYMWRMWRWRGWATGAGMKAWRASCMMRRKYAGKSKTSYKSMWMCWTRSWKYCYTKA 279

Qy        250 CAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGT 309
Db        280 RWTGYYCYRKGGMWGKRGRWYASKKYMWKRWWWCWARMYRYSTGTRASMWWRRWYYTMMM 339

Qy        310 ATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCA--GAAAGGTTCTGGTAATA 367
Db        340 KWWKYAWARAAWRWWAMWWAWRRACAAAATATAATTATTTATGGTACAATTCTTGTACTT 399

Qy        368 AAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGT 424
Db        400 TAGCAAATCTGGAGTTAGTTCATAGTCAAAGTCAGTTAATATTTCTTAGAGGAAAGT 456

Search completed: March 18, 2004, 20:20:09
Job time : 58.4971 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	36	7.9	413	14	US-10-184-644-156	Sequence 156, App
	2	36	7.9	413	14	US-10-184-634-156	Sequence 156, App
c	3	34.4	7.5	5184	14	US-10-247-671-55	Sequence 55, Appl
	4	34	7.5	12426	12	US-10-221-714A-204	Sequence 204, App
c	5	33.8	7.4	1093	12	US-10-424-599-43422	Sequence 43422, A
	6	33.8	7.4	1149	9	US-09-912-020-139	Sequence 139, App
	7	33.8	7.4	1149	12	US-10-282-122A-6441	Sequence 6441, Ap
	8	33.6	7.4	1299	12	US-10-282-122A-36109	Sequence 36109, A
	9	33.2	7.3	1371	12	US-10-282-122A-17259	Sequence 17259, A
	10	33	7.2	312	12	US-10-424-599-104721	Sequence 104721,
c	11	33	7.2	1194	15	US-10-369-493-32803	Sequence 32803, A
	12	33	7.2	1269	12	US-10-282-122A-35991	Sequence 35991, A
	13	33	7.2	1837	12	US-10-282-122A-35601	Sequence 35601, A
c	14	33	7.2	319630	15	US-10-398-221-7	Sequence 7, Appli
	15	33	7.2	3011208	15	US-10-398-221-2058	Sequence 2058, Ap
c	16	32.8	7.2	746	15	US-10-027-632-14807	Sequence 14807, A
c	17	32.8	7.2	746	15	US-10-027-632-14808	Sequence 14808, A
	18	32.8	7.2	2552	15	US-10-027-632-111939	Sequence 111939,
	19	32.6	7.1	351	9	US-09-983-965-4342	Sequence 4342, Ap
c	20	32.6	7.1	2000	15	US-10-260-238-1758	Sequence 1758, Ap
c	21	32.6	7.1	2250	15	US-10-369-493-46046	Sequence 46046, A
c	22	32.6	7.1	17419	14	US-10-239-676-100	Sequence 100, App
c	23	32.6	7.1	17419	14	US-10-311-455-1268	Sequence 1268, Ap
c	24	32.6	7.1	17419	14	US-10-240-453-112	Sequence 112, App
	25	32.4	7.1	1320	12	US-10-282-122A-10004	Sequence 10004, A
	26	32.4	7.1	2379	12	US-10-282-122A-9449	Sequence 9449, Ap
	27	32.2	7.1	2634	12	US-10-282-122A-35828	Sequence 35828, A
	28	32	7.0	2565	15	US-10-369-493-25168	Sequence 25168, A
c	29	31.8	7.0	409	9	US-09-864-761-2955	Sequence 2955, Ap
c	30	31.8	7.0	438	9	US-09-864-761-19736	Sequence 19736, A
	31	31.8	7.0	708	14	US-10-184-644-584	Sequence 584, App
	32	31.8	7.0	708	14	US-10-184-634-584	Sequence 584, App
	33	31.8	7.0	1860	12	US-10-282-122A-27280	Sequence 27280, A
	34	31.8	7.0	11138	10	US-09-960-870-5	Sequence 5, Appli
	35	31.8	7.0	11138	10	US-09-960-858-5	Sequence 5, Appli
	36	31.8	7.0	580073	14	US-10-205-220-1	Sequence 1, Appli
	37	31.8	7.0	640681	9	US-09-790-988-1	Sequence 1, Appli
	38	31.6	6.9	480	9	US-09-815-242-4053	Sequence 4053, Ap
	39	31.6	6.9	480	12	US-10-282-122A-7419	Sequence 7419, Ap
c	40	31.6	6.9	575	15	US-10-027-632-222766	Sequence 222766,
	41	31.6	6.9	2565	15	US-10-321-802-19	Sequence 19, Appl
c	42	31.4	6.9	867	15	US-10-027-632-167010	Sequence 167010,
c	43	31.4	6.9	867	15	US-10-027-632-167011	Sequence 167011,
c	44	31.4	6.9	867	15	US-10-027-632-167012	Sequence 167012,
c	45	31.4	6.9	1407	12	US-10-424-599-25067	Sequence 25067, A

ALIGNMENTS

RESULT 1
US-10-184-644-156
; Sequence 156, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 156
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-156

Query Match 7.9%; Score 36; DB 14; Length 413;
Best Local Similarity 11.6%; Pred. No. 2;
Matches 28; Conservative 93; Mismatches 121; Indels 0; Gaps 0;

```
Qy      147 TGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGC 206
            ::|:  ::  :: ::   ::: : | : :| ::::: |:||: :   : :     :||
Db      172 HDGKDYVKGSKKLRVGLLKMRNKSGGKKRRGSKRSRREASGGDQREGTREHLQERAKGGR 231

Qy      207 GGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAA 266
```

```
Db        232 RRKKSGRGQRIAEGRPSFQWTRVKNTHIPKGWARGGMGDATLDYDYALLELKRAHKKKYM 291

Qy        267 CCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTA 326

Db        292 ELGISPTIKKMPGGMIHFSGFDNDRADQLVYRFCSVSDESNDLLYQYCDAESGSTGSGVY 351

Qy        327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386

Db        352 LRLKDPDKKNWKRKIIAVYSGHQWVDVHGVQKDYNVAVRITPLKYAQICLWIHGNDANCA 411

Qy        387 TG 388

Db        412 YG 413


RESULT 2
US-10-184-634-156
; Sequence 156, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C217
;  CURRENT APPLICATION NUMBER: US/10/184,634
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 156
;   LENGTH: 413
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-634-156

  Query Match             7.9%;  Score 36;  DB 14;  Length 413;
  Best Local Similarity  11.6%;  Pred. No. 2;
  Matches   28;  Conservative   93;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        147 TGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGC 206

Db        172 HDGKDYVKGSKKLRVGLLKMRNKSGGKKRRGSKRSRREASGGDQREGTREHLQERAKGGR 231

Qy        207 GGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAA 266

Db        232 RRKKSGRGQRIAEGRPSFQWTRVKNTHIPKGWARGGMGDATLDYDYALLELKRAHKKKYM 291

Qy        267 CCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTA 326

Db        292 ELGISPTIKKMPGGMIHFSGFDNDRADQLVYRFCSVSDESNDLLYQYCDAESGSTGSGVY 351

Qy        327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386

Db        352 LRLKDPDKKNWKRKIIAVYSGHQWVDVHGVQKDYNVAVRITPLKYAQICLWIHGNDANCA 411

Qy        387 TG 388

Db        412 YG 413


RESULT 3
US-10-247-671-55/c
, Sequence 55, Application US/10247671
; Publication No. US20030194721A1
; GENERAL INFORMATION:
;  APPLICANT: Mikita, Thomas
;  APPLICANT:  Shiffman, Dov
;  APPLICANT:  Porter, Gordon, J.
;  APPLICANT:  Kaser, Matthew R.
;  TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
;  FILE REFERENCE: PA-0050 US
;  CURRENT APPLICATION NUMBER: US/10/247,671
;  CURRENT FILING DATE:  2002-09-18
;  PRIOR APPLICATION NUMBER: 60/323,784
;  PRIOR FILING DATE: 2001-09-19
;  NUMBER OF SEQ ID NOS: 186
;  SOFTWARE: PERL Program
; SEQ ID NO 55
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030194721A1 5185743CB1
US-10-247-671-55

  Query Match             7.5%;  Score 34.4;  DB 14;  Length 5184;
  Best Local Similarity  52.9%;  Pred. No. 23;
  Matches   74;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy         51 GATTGCAGCCGCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGA 110

Db       2489 GCTTGCAGCTGCTGAATAGTATTCTGAAGCTGACTCACACTATTTGGTGATGGAAAAATA 2430

Qy        111 ATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAG 2430

Db       2429 TTTGATGAAATCTGCTGCTGTAAAGTCTGTGCTTGTTCTTGTAGTGGTGACTGCTGCTGT 170

Qy        171 TGCTCAGTTACGGCAGGGAG 190

Db       2369 TGTTGTGATGCCTCAGTCAG 2350


RESULT 4
US-10-221-714A-204
; Sequence 204, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 204
;   LENGTH: 12426
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-204
```

```
  Query Match             7.5%;  Score 34;  DB 12;  Length 12426;
  Best Local Similarity   48.5%;  Pred. No. 50;
  Matches   94;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy         14 TGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATG 73
              |||| |||| || |    | | |    ||| | | |||  |   | ||    || |
Db       9253 TGTTTGTTATTGTGGTTTATAGATTTTAAGCGATTTGTATTTAAATGGTAGAGTTTTTTT 9312

Qy         74 ATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATC 133
              ||||   |  || |   | ||| ||      ||||   ||   | ||| ||  |||||||
Db       9313 ATTTTTTTTTTTGAAGTTTAATTTTTTTAAAATATTTTGTTTAGTGTTTGTTTAATC 9372

Qy        134 AGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCT 193
              ||      | |||  | | | | ||| |   |  |    | ||  |       | |   |
Db       9373 GGGGTAGAAATATTTTTAATGATAGGAATTTAATAATATTTAAGGAAGTTTTAGTAAAGT 9432

Qy        194 CAAAACTTTTGGCG 207
              |||  |||| | |
Db       9433 GAAATGTTTTTGGG 9446


RESULT 5
US-10-424-599-43422/c
; Sequence 43422, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 43422
;   LENGTH: 1093
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_139209C.1
US-10-424-599-43422

  Query Match             7.4%;  Score 33.8;  DB 12;  Length 1093;
  Best Local Similarity   58.4%;  Pred. No. 15;
  Matches   59;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        294 TGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAA 353
              ||| |||  | | |      ||  |||||| | | | | || | | ||  |||  | || |
Db        640 TGCAAACAGTTCGATGTCATCAGAAGGTGATAAAGTTTATCCCGAGAGTATTTGCAAGGA 581

Qy        354 AGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTC 394
              || |||| ||  |||  |   |||||    |||||| ||
Db        580 AGATTCTTGTGTTAAGTCTGCAGTTACAGCTAATGGTAGTC 540


RESULT 6
US-09-912-020-139
; Sequence 139, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Trawick, John
;  APPLICANT:  Forsyth, R. Allyn
;  APPLICANT:  Froelich, Jamie M.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001DV1
;  CURRENT APPLICATION NUMBER: US/09/912,020
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 139
;   LENGTH: 1149
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-912-020-139

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 1149;
  Best Local Similarity   56.9%;  Pred. No. 16;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        259 GATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAA 318
              ||| |||||||    ||||  | |||||  || |   || |||  |  | ||  |||
Db        442 GATAATAACCTGAATTATAGCGTTCAGGTCGGTAACACCCACGGAGGTAATACATCGTCT 501

Qy        319 GGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATA 367
              ||  |   |||| | | | |  | | |   |    || || ||||||||
Db        502 GGCACCAGTGGTTACAGTTCTCTTAATTATCGTGGAGCTTATGGTAATA 550


RESULT 7
US-10-282-122A-6441
; Sequence 6441, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6441
```

```
;   LENGTH: 1149
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-282-122A-6441

  Query Match             7.4%;  Score 33.8;  DB 12;  Length 1149;
  Best Local Similarity   56.9%;  Pred. No. 16;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        259 GATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAA 318
              ||| |||||||    ||||  | |||||  || |    || |||  |  | ||  |||
Db        442 GATAATAACCTGAATTATAGCGTTCAGGTCGGTAACACCCACGGAGGTAATACATCGTCT 501

Qy        319 GGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATA 367
              ||  |   |||| | | | | | | |    |    || || ||||||||
Db        502 GGCACCAGTGGTTACAGTTCTCTTAATTATCGTGGAGCTTATGGTAATA 550


RESULT 8
US-10-282-122A-36109
; Sequence 36109, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36109
;   LENGTH: 1299
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
US-10-282-122A-36109

  Query Match             7.4%;  Score 33.6;  DB 12;  Length 1299;
  Best Local Similarity   59.4%;  Pred. No. 19;
  Matches   57;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy        299 ACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTT 358
              | |||   |||| | ||||||  ||||||| | |  | |  | || | |  | || | |
Db        314 ATGATAATAGTACTGCGCAAGCAGCTTATGATCAAGCAGTCAGGAATCTAGATAAGGTTA 373

Qy        359 CTGGTAATAAAGCAAATATTACACAGTATGGTACTC 394
              || | |  |||||  || |   |||||||| |
Db        374 ACCGTCAAATTACAAATTATAAAACTTATGGTACGC 409


RESULT 9
US-10-282-122A-17259
; Sequence 17259, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17259
;   LENGTH: 1371
;   TYPE: DNA
;   ORGANISM: Clostridium difficile
US-10-282-122A-17259

  Query Match             7.3%;  Score 33.2;  DB 12;  Length 1371;
  Best Local Similarity   48.0%;  Pred. No. 26;
  Matches   95;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy         82 AATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCC 141
              |||  |||| || | |  || ||||  ||  || ||| ||||  ||| ||| |   |
Db        466 AATAAAGAACATGAGTGTGCTGTAATAGAGATGGGTATGTCTGGATTCAATGAAATAAAA 525

Qy        142 ATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTT 201
                | |  |  | | | | ||| | |||     ||  | |   ||| | |  || |
Db        526 TATTTAGTGGATATAGTAAATCCTAAAATTGCAGTCATATCGAATATAGGATTATCACAT 585

Qy        202 TTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGAT 261
              | ||    | |   | |||   |||         |||||||| ||    ||   | |
Db        586 GTAGAAAAACTAGGTTCACAAGAGGGTATTTTGAAAGCAAAGATGGAGATTACTTCAAAC 645
```

```
Qy        262 TATAACCTTGCATATATT 279
              | | |     || ||| |
Db        646 TTTGATGAAACAAATACT 663
```

RESULT 10
US-10-424-599-104721
; Sequence 104721, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 104721
; LENGTH: 312
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_65579C.1
US-10-424-599-104721

Query Match 7.2%; Score 33; DB 12; Length 312;
Best Local Similarity 52.6%; Pred. No. 13;
Matches 72; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        267 CCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTA 326
              ||·||| |  | |||| |  |   | |||||   | |||  |  |||| |||| ||  |
Db        129 CCATGCTTTGAATGATAAAACTTACCGTGCCCGTGCTGCATGGTTTTCCCAAGCGGCAAA 188

Qy        327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386
                || |    |  ||||  ||| |   ||  ||||||  |||  |    |    || |
Db        189 AAAAAAAAAAAAGGGGATTGGCCATACGAGTCATGGTAACCAAGAGAGACATGATGAGCA 248

Qy        387 TGGTACTCAAAAAACGG 403
              | ||  | | | || |
Db        249 CGATAAACGATATACTG 265
```

RESULT 11
US-10-369-493-32803/c
; Sequence 32803, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 32803
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Chloroflexus aurantiacus
US-10-369-493-32803

Query Match 7.2%; Score 33; DB 15; Length 1194;
Best Local Similarity 48.6%; Pred. No. 28;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```
Qy        167 ATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTA 226
              | | ||  | |||     ||| ||  | | ||  ||  | |  |||     |||  |||||
Db        737 ACACTGTCCGGTTCGACCAGCGAATCACCAACATTCAGCCCTTTGGCAAGCAGAAAGGTG 678

Qy        227 GTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGG 286
              | || |     | |  | ||| |||       ||   || ||       | | ||||| |
Db        677 TTGGCCATACCACCACCAATGAGCAGTGCATCGACCTTACCCAGCAGATTTTCGATCACG 618

Qy        287 CGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTA 346
              | |    || ||  |||    | |   |  | |  | ||| |     ||   |  |||||
Db        617 CCGATCTTGTCACTGATTTTCGCACCACCGATGACGGTTACGAATGGACGCTGCGGATTA 558

Qy        347 TCCAG 351
              | |||
Db        557 TTCAG 553
```

RESULT 12
US-10-282-122A-35991
; Sequence 35991, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35991
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Streptococcus mutans
US-10-282-122A-35991

Query Match 7.2%; Score 33; DB 12; Length 1269;
Best Local Similarity 49.2%; Pred. No. 29;
Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```
Qy        236 GGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTG 295
```

```
Db        413 GAGCAAAGATTCAACAGACAGGTGGTGATATTACAGCCACTGCCGATCGTTTAAAGGGTG 472
Qy        296 CCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAG 355
Db        473 CTAACATTTACATGGATTTTCCAAGTGTTGGCGCGACTCAGAATCTTATGATGGCTGCTA 532
Qy        356 GTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAG 412
Db        533 CTTTAGCTGATGGAACAACGATTATCGAAAATGCAGCACGTGAACCAGAAATCGTTG 589
```

RESULT 13
US-10-282-122A-35601
; Sequence 35601, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35601
; LENGTH: 1837
; TYPE: DNA
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35601

Query Match 7.2%; Score 33; DB 12; Length 1837;
Best Local Similarity 47.8%; Pred. No. 35;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy        197 AACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAG 256
Db       1449 AAATAGAGACTGATACTGGTAAACGATTAATTATTAAACTTGAAACAATTAGTGAGCCGG 1508
Qy        257 GAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGC 316
Db       1509 ATGAAAATGGTTATCGAACAATTTATTATGTTATGAATGGTCAAGCACGTCGTATTTCTA 1568
Qy        317 AAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATA 376
Db       1569 TAAAAGATGAAAATATTAAAACAAATACGAACTTGAAACCTAAAGCTGACAAATCAAATC 1628
Qy        377 TTACACAGTATGGTACTCAAA 397
Db       1629 CTAGTCATATTGGTGCTCAAA 1649
```

RESULT 14
US-10-398-221-7/c
; Sequence 7, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 319630
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-7

Query Match 7.2%; Score 33; DB 15; Length 319630;
Best Local Similarity 52.6%; Pred. No. 5.6e+02;
Matches 72; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        249 CCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAG 308
Db     285383 CCCTTCACGAGTTCATAGCCTTTCGTATTTCGATATATAGTTTTGAAATATAAATCACCA 285324
Qy        309 TATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAA 368
Db     285323 TATTTCAAAACTATTGTATGATAATTCGGCTATTAAAGTCAAGTTATTTTGAAATATAGG 285264
Qy        369 AGCAAATATTACACAGT 385
Db     285263 TGCTAATATTTCAAAAT 285247
```

RESULT 15
US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208

```
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2058

  Query Match             7.2%;  Score 33;  DB 15;  Length 3011208;
  Best Local Similarity   52.6%;  Pred. No. 9.5e+02;
  Matches   72;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        249 CCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAG 308
              ||   || ||| | ||| |||| | ||| | |||     |    |    |   ||  |
Db    1832040 CCCTTCACGAGTTCATAGCCTTTCGTATTTCGATATATAGTTTTGAAATATAAATCACCA 1832099

Qy        309 TATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAA 368
              ||||||  ||       |||| |||| | || || |    || ||  |  ||    ||
Db    1832100 TATTTCAAAACTATTGTATGATAATTCGGCTATTAAAGTCAAGTTATTTTGAAATATAGG 1832159

Qy        369 AGCAAATATTACACAGT 385
              || |||||| || | |
Db    1832160 TGCTAATATTTCAAAAT 1832176


Search completed: March 19, 2004, 03:48:55
Job time : 257.734 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-4	Sequence 4, Appli
2	332.8	73.0	456	23	US-09-543-407-2	Sequence 2, Appli
3	290.2	63.6	561	17	US-09-252-691-171	Sequence 171, App
4	290.2	63.6	561	17	US-09-252-691C-171	Sequence 171, App
5	290.2	63.6	561	51	US-10-417-886-171	Sequence 171, App
6	100	21.9	100	45	US-10-146-492B-80	Sequence 80, Appl
7	100	21.9	100	45	US-10-146-492B-81	Sequence 81, Appl
8	100	21.9	100	45	US-10-146-492B-82	Sequence 82, Appl
c 9	39.6	8.7	7218	8	US-08-466-194-14	Sequence 14, Appl
10	38.2	8.4	391	18	US-09-332-782-15700	Sequence 15700, A
11	38.2	8.4	391	21	US-09-515-694-15700	Sequence 15700, A
12	38.2	8.4	4860	51	US-10-431-652-3221	Sequence 3221, Ap
c 13	37.8	8.3	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
14	37.4	8.2	456	23	US-09-543-407-13	Sequence 13, Appl
c 15	37.4	8.2	68667	39	US-09-949-016-17017	Sequence 17017, A
c 16	37.4	8.2	76668	103	US-60-466-412-82785	Sequence 82785, A
c 17	36.8	8.1	573	43	US-10-021-323-3919	Sequence 3919, Ap
c 18	36.8	8.1	573	82	US-60-255-619-3919	Sequence 3919, Ap
19	36.6	8.0	456	23	US-09-543-407-25	Sequence 25, Appl
c 20	36.4	8.0	396071	103	US-60-465-241-51735	Sequence 51735, A
21	36.2	7.9	42918	103	US-60-466-412-85430	Sequence 85430, A
22	36.2	7.9	2105303	39	US-09-947-916-273	Sequence 273, App
23	36	7.9	413	46	US-10-176-912-156	Sequence 156, App
24	36	7.9	413	46	US-10-179-524-156	Sequence 156, App
25	36	7.9	413	46	US-10-184-634-156	Sequence 156, App
26	36	7.9	413	46	US-10-184-644-156	Sequence 156, App
c 27	36	7.9	395143	103	US-60-466-412-87989	Sequence 87989, A
c 28	36	7.9	395143	106	US-60-499-964-5135	Sequence 5135, Ap
29	35.8	7.9	85690	29	US-09-692-412-72	Sequence 72, Appl
30	35.8	7.9	85690	33	US-09-803-736-1052	Sequence 1052, Ap
31	35.8	7.9	85690	55	US-10-746-294-72	Sequence 72, Appl
32	35.6	7.8	9493	70	US-60-144-351-3079	Sequence 3079, Ap
33	35.6	7.8	12178	26	US-09-620-392-28400	Sequence 28400, A
34	35.6	7.8	12178	30	US-09-702-134-10097	Sequence 10097, A
35	35.6	7.8	12178	33	US-09-815-264-80300	Sequence 80300, A
c 36	35.6	7.8	13036	26	US-09-620-392-47103	Sequence 47103, A
c 37	35.6	7.8	13036	30	US-09-702-134-22694	Sequence 22694, A
c 38	35.6	7.8	13036	33	US-09-815-264-64696	Sequence 64696, A
39	35.6	7.8	32948	26	US-09-620-392-40033	Sequence 40033, A
40	35.6	7.8	32948	30	US-09-702-134-20486	Sequence 20486, A
41	35.6	7.8	32948	33	US-09-815-264-66935	Sequence 66935, A
c 42	35.6	7.8	59719	39	US-09-949-016-15616	Sequence 15616, A
c 43	35.6	7.8	163183	107	US-60-500-337-19909	Sequence 19909, A
44	35.4	7.8	610	44	US-10-098-754-5821	Sequence 5821, Ap
45	35.4	7.8	837105	39	US-09-948-124-76	Sequence 76, Appl

ALIGNMENTS

```
RESULT 1
US-09-543-407-4
; Sequence 4, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-4

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 9e-125;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456


RESULT 2
US-09-543-407-2
; Sequence 2, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-2

  Query Match           73.0%;  Score 332.8;  DB 23;  Length 456;
  Best Local Similarity  83.1%;  Pred. No. 4.7e-88;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  |||||||  | |||  ||||||||||||  || ||||||||||||| || |||
Db         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
```

```
              ||||||||||||||||||  |||||  ||||||||||  ||  ||  ||||||||||      ||
Db        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||  ||||  |||  ||||||||  |||  |  |||  ||  |  ||||||||    |  |   ||  |||||
Db        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||   |  ||||||  ||||   ||||||||  ||||  ||||||||  ||   ||||||  |||||||
Db        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||  ||||||  |  |||||  |||||||  |||    ||||||||||||||||||
Db        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||  |||||  ||  ||||||||  |||||  |||||  ||  |||||  |||  ||||||||||||  |
Db        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||  ||||||||||||||  ||  |||||  |||
Db        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456


RESULT 3
US-09-252-691-171
; Sequence 171, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-171

  Query Match             63.6%;  Score 290.2;  DB 17;  Length 561;
  Best Local Similarity   77.4%;  Pred. No. 2.5e-75;
  Matches  352;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||  |||  ||||||||||| |||  |||||||||||||||||  |  |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||  | ||  ||||||||  | |||  ||||||||  || |||  ||||||||||  |||||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
               ||||||  ||||||||||||||  |||||||||    ||  |    ||||  |  ||  |  ||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||  ||||  ||  ||  |||  |  ||  |  |||  ||  |  ||  ||  |  |  ||||||  ||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  ||||  |||  ||||     ||||||||||  |||||  ||||||  |  ||||   |  ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||  ||||||  |||  ||  ||||  |||  ||  ||  ||  ||||||||||||||||||  ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              |||||  |  ||  |||||  ||  ||||||||||  ||  |||||  |||  |||||||  ||||||
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||  |  ||||||||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 4
US-09-252-691C-171
; Sequence 171, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-171

  Query Match             63.6%;  Score 290.2;  DB 17;  Length 561;
  Best Local Similarity   77.4%;  Pred. No. 2.5e-75;
  Matches  352;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||  |||  ||||||||||| |||  |||||||||||||||||  |  |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||  | ||  ||||||||  | |||  ||||||||  || |||  ||||||||||  |||||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
               ||||||  ||||||||||||||  |||||||||    ||  |    ||||  |  ||  |  ||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||  ||||  ||  ||  |||  |  ||  |  |||  ||  |  ||  ||  |  |  ||||||  ||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  ||||  |||  ||||     ||||||||||  |||||  ||||||  |  ||||   |  ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||  ||||||  |||  ||  ||||  |||  ||  ||  ||  ||||||||||||||||||  ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              |||||  |  ||  |||||  ||  ||||||||||  ||  |||||  |||  |||||||  ||||||
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTA 455
              ||||||||||||||||||||||  |  ||||||||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 5
US-10-417-886-171
; Sequence 171, Application US/10417886
; GENERAL INFORMATION:
;  APPLICANT: Keith G. Weinstock et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.135
;  CURRENT APPLICATION NUMBER: US/10/417,886
;  CURRENT FILING DATE:  2003-04-17
```

```
;  PRIOR APPLICATION NUMBER: US/09/252,691C
;  PRIOR FILING DATE: 1999-02-18
;  PRIOR APPLICATION NUMBER: US 60/094,145
;  PRIOR FILING DATE: 1998-07-24
;  PRIOR APPLICATION NUMBER: US 60/074,787
;  PRIOR FILING DATE: 1998-02-18
;  NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-10-417-886-171

  Query Match           63.6%;  Score 290.2;  DB 51;  Length 561;
  Best Local Similarity  77.4%;  Pred. No. 2.5e-75;
  Matches  352;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||  ||| ||||||||||| ||| ||||||||||||||||| ||| |  |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||  | ||  |||||||| | ||| |||||||| || ||| |||||||||| |||||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||  |||||||||||||| |||||||||    ||  |    |||| | || |  ||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || ||||  || || ||| |  || | ||| || | || || ||  | |  ||||| ||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  |||| ||| ||||    |||||||||| ||||| |||||| | ||||  | ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||| ||||||  ||| || |||| ||| || || || ||||||||||||||||| ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||| |  || ||||| || |||||||||||| || ||||| ||| ||||||| ||||||
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTA 455
              |||||||||||||||||||||| |  ||||||||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 6
US-10-146-492B-80
; Sequence 80, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 80
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 39-58
US-10-146-492B-80

  Query Match           21.9%;  Score 100;  DB 45;  Length 100;
  Best Local Similarity  100.0%;  Pred. No. 6.7e-19;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        266 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 60

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 365
              |||||||||||||||||||||||||||||||||||||||
Db         61 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 100


RESULT 7
US-10-146-492B-81
; Sequence 81, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 81
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 38-57
US-10-146-492B-81

  Query Match           21.9%;  Score 100;  DB 45;  Length 100;
  Best Local Similarity  100.0%;  Pred. No. 6.7e-19;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        208 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 60

Qy        268 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 307
              |||||||||||||||||||||||||||||||||||||||
Db         61 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 100


RESULT 8
US-10-146-492B-82
; Sequence 82, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 82
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 57-76
US-10-146-492B-82

  Query Match           21.9%;  Score 100;  DB 45;  Length 100;
  Best Local Similarity  100.0%;  Pred. No. 6.7e-19;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        123 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 60

Qy        183 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 222
              |||||||||||||||||||||||||||||||||||||||
Db         61 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 100


RESULT 9
US-08-466-194-14/c
; Sequence 14, Application US/08466194
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, NW, Suite 500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20007
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,194
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/232,463
;      FILING DATE:  22-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/935,313
;      FILING DATE:  26-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/201 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-466-194-14

  Query Match             8.7%;  Score 39.6;  DB 8;  Length 7218;
  Best Local Similarity   10.5%;  Pred. No. 2.8;
  Matches   39;  Conservative  166;  Mismatches  165;  Indels    0;  Gaps    0;

Qy         64 GCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCT 123
              |||  |  | | ||    ||  | | || ||          |||  |  |  || |
Db       1523 GCATCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTACCTATCTATGCAA 1464

Qy        124 TCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGG 183
                | ||||  ||  ||    | |||||  ::  ::::  :: :: :: :   ::  : ::
Db       1463 GTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1404

Qy        184 CAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAG 243
               ::::::   ::::     :: ::  :  : : ::::::: :: :: ::  ::: ::::
Db       1403 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1344

Qy        244 ATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGAT 303
              :  ::  ::: :::::  : ::    : : : :  ::  ::: ::: :: :  :: ::
Db       1343 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1284

Qy        304 GCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGT 363
              :  :: :   : :::: :   : :: :: :  : :: ::  :   :::::::    ::
Db       1283 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1224

Qy        364 AATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAG 423
              :: :::: ::: :  : : :: : :: :   :::::: :: ::  : :: : ::::: ::
Db       1223 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1164

Qy        424 TCGCAAATGG 433
                : ::: ::
Db       1163 RRRRRRRRR 1154


RESULT 10
US-09-332-782-15700
; Sequence 15700, Application US/09332782
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-754CON1
;  CURRENT APPLICATION NUMBER: US/09/332,782
;  CURRENT FILING DATE: 1999-06-14
;  EARLIER APPLICATION NUMBER: US 09/181,430
;  EARLIER FILING DATE: 1998-10-28
;  NUMBER OF SEQ ID NOS: 21027
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15700
;   LENGTH: 391
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-332-782-15700

  Query Match             8.4%;  Score 38.2;  DB 18;  Length 391;
  Best Local Similarity   58.3%;  Pred. No. 2.7;
  Matches   67;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        309 TATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAA 368
              | |  |||||   |||  |||| || |||   ||  ||| | | ||   ||||| |||
Db        277 TTTCGCGCAACAAGCTGGTGGTCATGCTGAACTGGATATTCCGCAATAATCTGGAAATGT 336

Qy        369 AGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAG 423
              ||| || | | || | |||| |    | ||   | || || |   ||| | || |
Db        337 AGCCAAGACTGCAAATTATGATGACAATAATCAGACACTTCTGTAGCATATACTG 391


RESULT 11
US-09-515-694-15700
; Sequence 15700, Application US/09515694
; GENERAL INFORMATION:
;  APPLICANT: Arterburn, Matthew
;  APPLICANT: Asghari, Vida
;  APPLICANT: Blun, Linda
;  APPLICANT: Cheung, Patrick
;  APPLICANT: Damavandi, Simin
;  APPLICANT: Dickson, Mark
;  APPLICANT: Drake, Jim
;  APPLICANT: Drmanac, Radoje
;  APPLICANT: Engleman, Carrie
;  APPLICANT: Faulkner, Brandy
;  APPLICANT: Fox, Melvin
;  APPLICANT: Garcia, Veronica
;  APPLICANT: Giedt, Gretchen
;  APPLICANT: Jomek, Leni
;  APPLICANT: Jones, Lee
;  APPLICANT: Kita, David
```

```
;  APPLICANT: Kofler, Janette
;  APPLICANT: Labat, Ivan
;  APPLICANT: Lee, Won-Jae
;  APPLICANT: Lomelli, Michelle
;  APPLICANT: Nguyen, Hong
;  APPLICANT: Nguyen, Linh
;  APPLICANT: Nguyen, Lynne
;  APPLICANT: Nguyen, Phuong
;  APPLICANT: Nogra, Margie
;  APPLICANT: Ojeda, Jesse
;  APPLICANT: Palencia, Servando
;  APPLICANT: Raisi, Fariba
;  APPLICANT: Randhwa, Gurpreet
;  APPLICANT: Sahourieh, Hannah
;  APPLICANT: Sidhu, Navjiwan
;  APPLICANT: Smith, Benjamin
;  APPLICANT: Smythe, Ashleigh
;  APPLICANT: Tkach, Joe
;  APPLICANT: Tulpule, Mukul
;  APPLICANT: Verna, Ron
;  APPLICANT: Wachter, Adam
;  APPLICANT: Wu, James
;  APPLICANT: Yim, Kenneth
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 754CIP
;  CURRENT APPLICATION NUMBER: US/09/515,694
;  CURRENT FILING DATE: 2000-02-29
;  EARLIER APPLICATION NUMBER: 09/332,782
;  EARLIER FILING DATE: 1999-06-14
;  EARLIER APPLICATION NUMBER: 09/181,430
;  EARLIER FILING DATE: 1998-10-28
;  NUMBER OF SEQ ID NOS: 21027
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15700
;   LENGTH: 391
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-515-694-15700

  Query Match             8.4%;  Score 38.2;  DB 21;  Length 391;
  Best Local Similarity   58.3%;  Pred. No. 2.7;
  Matches   67;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        309 TATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAA 368
              | |  |||||   |||  |||| || |||   ||  ||| | | ||   ||||| |||
Db        277 TTTCGCGCAACAAGCTGGTGGTCATGCTGAACTGGATATTCCGCAATAATCTGGAAATGT 336

Qy        369 AGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAG 423
              ||| || | | || | |||| |   | ||   | || || |   ||| | || |
Db        337 AGCCAAGACTGCAAATTATGATGACAATAATCAGACACTTCTGTAGCATATACTG 391


RESULT 12
US-10-431-652-3221
; Sequence 3221, Application US/10431652
; GENERAL INFORMATION:
;  APPLICANT: Breton, Gary L.
;  APPLICANT:  Bush, David
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION:  BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH03-08
;  CURRENT APPLICATION NUMBER: US/10/431,652
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: US 09/328,352
;  PRIOR FILING DATE: 1999-06-04
;  PRIOR APPLICATION NUMBER: US 60/088,701
;  PRIOR FILING DATE: 1998-06-09
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3221
;   LENGTH: 4860
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-10-431-652-3221

  Query Match             8.4%;  Score 38.2;  DB 51;  Length 4860;
  Best Local Similarity   45.9%;  Pred. No. 6.3;
  Matches  130;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy        106 AATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAAT 165
              |||    | |||| |  | |  || | | |||||     | |  |||||| |   || |
Db       3445 AATACTATTGGTAAATTATTAGGTACTGATGCAGCAGATAATCTTCAAGCAGATGCTGAG 3504

Qy        166 AATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGT 225
              | |    ||   |   |   |||||||    |     |    | || ||  ||  | | |
Db       3505 ATTTCAACTATTTATGGTTTGGGAGGCAATGATACGATTCAAGGTGGTGTACAGAATGAT 3564

Qy        226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
              |     | | |   |  ||| | |||  || || |||    | | | |  | | |
Db       3565 TATCTGTATGGTGATGATGGAGACGATACATTAGTTTCTAATACAGGCTCTGATTATTTG 3624

Qy        286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
              ||  ||||     ||| ||   |     | || | | ||   |||       | | |
Db       3625 TATGGTGGTGCAGGTAATGACACATTGATTTATGGCGGTAATTCCAATGTGTATACGGCT 3684

Qy        346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATG 388
              | ||         |||| ||| |  |  |||||| | |  | |
Db       3685 TTACAAGGTCAAGCTGGGAATGACACTTATATTATAGATAAAG 3727


RESULT 13
US-09-887-272A-5263/c
; Sequence 5263, Application US/09887272A
; GENERAL INFORMATION:
;  APPLICANT: Hou, Yu-Ming
;  APPLICANT:  Quan, Sheng
;  APPLICANT:  Chang, Hur-Song
;  APPLICANT:  Zhu, Tong
;  APPLICANT:  Whitham, Steve
;  APPLICANT:  Goff, Steve
;  APPLICANT:  Glazebrook, Jane
;  APPLICANT:  Chen, Wenquiong
;  APPLICANT:  Katagiri, Fumiaki
;  APPLICANT:  Xie, Zhiyi
;  APPLICANT:  Tao, Yi
;  APPLICANT:  Zou, Guangzhou
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: PLANT GENES INVOLVED IN DEFENSE AGAINST
;  TITLE OF INVENTION:  PATHOGENS
;  FILE REFERENCE: 1360.003US2
;  CURRENT APPLICATION NUMBER: US/09/887,272A
;  CURRENT FILING DATE:  2001-06-23
;  PRIOR APPLICATION NUMBER: 60/213,634
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/214,926
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/261,320
;  PRIOR FILING DATE: 2001-01-12
;  PRIOR APPLICATION NUMBER: 60/264,353
;  PRIOR FILING DATE: 2001-01-26
;  PRIOR APPLICATION NUMBER: 60/273,879
;  PRIOR FILING DATE: 2001-03-07
;  PRIOR APPLICATION NUMBER: 09/887,271
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 6813
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5263
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-887-272A-5263

  Query Match             8.3%;  Score 37.8;  DB 36;  Length 2000;
  Best Local Similarity   13.1%;  Pred. No. 6.2;
  Matches   36;  Conservative  121;  Mismatches  118;  Indels    0;  Gaps    0;
```

```
Qy          21 TATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATGATTTAGC 80
               :| |:::: :|  :: |  : ::: :|   : :      : ::  | ::  : :   :
Db         950 WAAGRWRWRMAWCWYCCMWKWWKMTSCMWWKYWRTWSCWYTMWWGAMRYAYYAMRRRRWT 891

Qy          81 TAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGC 140
               : : :  :|:| : :|      :| : :||  | :::  :| : : :|:|||:   ::
Db         890 YKWSWRRMYWTMTKWAWTWMTCMCMAKWYMATGWATWMWWRYTMYTYCYAMTCAKCKYKM 831

Qy         141 CATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACT 200
                :| ::|   ||:: |: ::  :| ::: ::::: | :::  :|:: |: :    : : :
Db         830 AMTKWWTTWACAWRATSWRWRAMAGMRWKRYKMKRAYWWRWWRCWKAGWARWMKSRYRWK 771

Qy         201 TTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGA 260
               :::   : ::: : :  :::::: :   : : :   : :|::       |    :|::|::
Db         770 WKKYATRYYWKMWAMTWWWSWRRWKSYRMWSGMGRMRWSAWRYCSRMKCAKTKYASSARW 711

Qy         261 TTATAACCTTGCATATATTGATCAGGCGGGCAGTG 295
               |::  ::: :: :: :  :|:|:: :: ::|: |:
Db         710 TKRAKRSYRYRRRWYWKRKGWTYRYRYWRSCRMTR 676


RESULT 14
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             8.2%;  Score 37.4;  DB 23;  Length 456;
  Best Local Similarity   51.5%;  Pred. No. 5;
  Matches   86;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
               |||   |||  ||  ||  ||||  |||  | |    || |||||  ||  || ||||
Db         208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Qy         337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
                 | |  ||||| ||||   | ||||| || |  ||    ||||       || ||
Db         268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327

Qy         397 AAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGGCTATTCGCGT 443
               ||  | |  |||   ||       | |    ||  ||| ||  || ||
Db         328 AACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGT 374


RESULT 15
US-09-949-016-17017/c
; Sequence 17017, Application US/09949016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17017
;   LENGTH: 68667
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(68667)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17017

  Query Match             8.2%;  Score 37.4;  DB 39;  Length 68667;
  Best Local Similarity   58.6%;  Pred. No. 27;
  Matches   65;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386
               || | || ||| ||| | | |||  ||||  |  || ||||||  | | || | | | ||
Db       17497 TGCTCATTCTGTGATAAATTTCCTTAAAGTGTAAGGGAATAAATAATAAATAAAATATTA 17438

Qy         387 TGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGGCTAT 437
               || ||    || | | |||  |||  ||    |  | |  |||||  || |
Db       17437 TGATATGAGAATAGCAGCATGTGTGCTGTGATGTTACTTCCAAATTACTTT 17387


Search completed: March 19, 2004, 03:01:45
Job time : 2654.77 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	37.4	8.2	65558	6	US-10-765-790-103	Sequence 103, App
	2	35	7.7	675	5	US-09-741-873C-3	Sequence 3, Appli
c	3	34.4	7.5	519599	6	US-10-765-790-73	Sequence 73, Appl
	4	34.2	7.5	887	6	US-10-767-701-9911	Sequence 9911, Ap
	5	33.8	7.4	1149	6	US-10-771-241-139	Sequence 139, App
c	6	33.4	7.3	38719	6	US-10-765-790-43	Sequence 43, Appl
c	7	32.2	7.1	261922	6	US-10-767-471-10653	Sequence 10653, A
	8	31.6	6.9	512	6	US-10-767-701-31198	Sequence 31198, A
	9	31.4	6.9	525	6	US-10-417-884A-2325	Sequence 2325, Ap
	10	31.4	6.9	1245	7	US-60-546-745-36	Sequence 36, Appl
	11	31.4	6.9	53332	6	US-10-786-065-3	Sequence 3, Appli
c	12	30.8	6.8	201	6	US-10-767-471-8836	Sequence 8836, Ap
c	13	30.8	6.8	201	6	US-10-767-471-34493	Sequence 34493, A
c	14	30.8	6.8	1227	6	US-10-767-471-501	Sequence 501, App
c	15	30.8	6.8	76573	6	US-10-767-471-10718	Sequence 10718, A
c	16	30.4	6.7	1227	6	US-10-417-884A-2544	Sequence 2544, Ap
	17	30.2	6.6	3529	6	US-10-779-543-9918	Sequence 9918, Ap
c	18	30.2	6.6	1790242	6	US-10-767-471-10805	Sequence 10805, A
c	19	29.8	6.5	600	7	US-60-545-213-3515	Sequence 3515, Ap
c	20	29.8	6.5	600	7	US-60-545-213-7787	Sequence 7787, Ap
c	21	29.6	6.5	201	6	US-10-767-471-8837	Sequence 8837, Ap
c	22	29.6	6.5	201	6	US-10-767-471-34439	Sequence 34439, A
	23	29.6	6.5	40000	6	US-10-767-471-10911	Sequence 10911, A
c	24	29.6	6.5	122673	6	US-10-765-790-33	Sequence 33, Appl
	25	29.4	6.4	2806	6	US-10-021-698A-2111	Sequence 2111, Ap
	26	29.4	6.4	74677	6	US-10-767-471-10891	Sequence 10891, A
	27	29.2	6.4	1080	6	US-10-451-467A-619	Sequence 619, App
c	28	29.2	6.4	134134	6	US-10-767-471-10743	Sequence 10743, A
c	29	29.2	6.4	189268	6	US-10-021-698A-697	Sequence 697, App
	30	29.2	6.4	290892	6	US-10-767-471-10809	Sequence 10809, A
c	31	29	6.4	447	6	US-10-767-701-18004	Sequence 18004, A
	32	29	6.4	532	6	US-10-767-701-17618	Sequence 17618, A
c	33	29	6.4	13948	6	US-10-767-471-10858	Sequence 10858, A
	34	29	6.4	16201	6	US-10-767-471-10705	Sequence 10705, A
	35	29	6.4	45268	7	US-60-548-091-5603	Sequence 5603, Ap
c	36	28.8	6.3	1829	6	US-10-775-169-326	Sequence 326, App
	37	28.8	6.3	1914	6	US-10-767-701-13618	Sequence 13618, A
	38	28.8	6.3	20767	6	US-10-100-683-12012	Sequence 12012, A
	39	28.6	6.3	408	6	US-10-417-884A-793	Sequence 793, App
c	40	28.6	6.3	1056	6	US-10-781-014-285	Sequence 285, App
c	41	28.6	6.3	1056	6	US-10-781-014-449	Sequence 449, App
	42	28.6	6.3	78785	5	US-09-978-167A-3	Sequence 3, Appli
	43	28.6	6.3	100374	7	US-60-548-091-5675	Sequence 5675, Ap
c	44	28.6	6.3	189817	6	US-10-767-471-10675	Sequence 10675, A
c	45	28.4	6.2	477	6	US-10-779-543-18788	Sequence 18788, A

ALIGNMENTS

RESULT 1
US-10-765-790-103/c
; Sequence 103, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 103
; LENGTH: 65558
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-103

Query Match 8.2%; Score 37.4; DB 6; Length 65558;
Best Local Similarity 58.6%; Pred. No. 0.21;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy      327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386
            || | || ||| ||| | | |||  ||||  |  || ||||||  | | || | | | ||
Db    16503 TGCTCATTCTGTGATAAATTTCCTTAAAGTGTAAGGGAATAAATAATAAATAAAATATTA 16444

Qy      387 TGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGGCTAT 437
            || ||    || | | |||  |||  ||    |  | |  |||||  || |
Db    16443 TGATATGAGAATAGCAGCATGTGTGCTGTGATGTTACTTCCAAATTACTTT 16393
```

RESULT 2
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1

```
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-09-741-873C-3

  Query Match             7.7%;  Score 35;  DB 5;  Length 675;
  Best Local Similarity   100.0%;  Pred. No. 0.24;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        416 AGAGACAGTCGCAAATGGCTATTCGCGTGACACAA 450
              |||||||||||||||||||||||||||||||||||
Db          1 AGAGACAGTCGCAAATGGCTATTCGCGTGACACAA 35


RESULT 3
US-10-765-790-73/c
; Sequence 73, Application US/10765790
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73
;   LENGTH: 519599
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-73

  Query Match             7.5%;  Score 34.4;  DB 6;  Length 519599;
  Best Local Similarity   60.9%;  Pred. No. 3.4;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        263 ATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTG 322
              ||||  ||| || |  |||||||  |    | || | |  || ||| |||| | || | |
Db     332273 ATAAAATTGAATTTTGTGATCAGTAGCCATGGGCTATCTCTGACAGCATTTGGAAAAGAG 332214

Qy        323 CTTATGGTAATACTGCGATGATTATCCAGAAA 354
              ||   | | |||| |   |  || | ||| |||
Db     332213 CTATAGTTGATACAGAAAATATAACCCAAAAA 332182


RESULT 4
US-10-767-701-9911
; Sequence 9911, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 9911
;   LENGTH: 887
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS40598_1
US-10-767-701-9911

  Query Match             7.5%;  Score 34.2;  DB 6;  Length 887;
  Best Local Similarity   53.3%;  Pred. No. 0.47;
  Matches   72;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        203 TGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATT 262
              || |  |  || |  |||  |||  ||  |   ||| |||  | ||   | |   |||
Db        433 TGTCAATGATTCCTAAAGGTGGTGATACAATTTGGGGAAATCTGGATTGGTCTCCAGAAG 492

Qy        263 ATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTG 322
              ||  |||||  || ||  | | ||  |  ||   |||| ||| ||    ||||  | | |
Db        493 ATGGCCTTGAATGTAAAGCTAAGAAGCACAAAACCAATGATACCGAGGTTTCTAAGGATA 552

Qy        323 CTTATGGTAATACTG 337
              ||||  | | ||
Db        553 GCAATGGGGAAAATG 567


RESULT 5
US-10-771-241-139
; Sequence 139, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 139
;   LENGTH: 1149
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-139

  Query Match             7.4%;  Score 33.8;  DB 6;  Length 1149;
  Best Local Similarity   56.9%;  Pred. No. 0.67;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        259 GATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAA 318
              |||  |||||||    ||||  | |||||  || |   || |||  |  | ||  |||
Db        442 GATAATAACCTGAATTATAGCGTTCAGGTCGGTAACACCCACGGAGGTAATACATCGTCT 501

Qy        319 GGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATA 367
              ||  |   |||| | | | |  | | |   |    || || ||||||||
Db        502 GGCACCAGTGGTTACAGTTCTCTTAATTATCGTGGAGCTTATGGTAATA 550


RESULT 6
```

```
US-10-765-790-43/c
; Sequence 43, Application US/10765790
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43
;   LENGTH: 38719
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-43

  Query Match             7.3%;  Score 33.4;  DB 6;  Length 38719;
  Best Local Similarity   51.7%;  Pred. No. 2.9;
  Matches   76;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         17 TATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATGATT 76
              | || ||||| |  | ||  |||     | | |     |   ||  | |       ||| |
Db      11094 TGTTGATGATATGGAGAAATCTGAACCCTCATACACTGCTAGTGGAAATGTAAAATGGTG 11035

Qy         77 TAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGG 136
              |||||| || |||| |||| || || ||     || ||| ||| | |  | |||  |
Db      11034 TAGCTACTTTAGAAAATAAGTTAGCAGTTCCACAAATGATTAAATATAGAGTTACCCTAT 10975

Qy        137 CAGCCATAATTGGTCAAGCTGGGACTA 163
              | |||  | |    |  || |   ||
Db      10974 GATCCAGCAATTACAATCCTAGTGATA 10948


RESULT 7
US-10-767-471-10653/c
; Sequence 10653, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10653
;   LENGTH: 261922
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(261922)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10653

  Query Match             7.1%;  Score 32.2;  DB 6;  Length 261922;
  Best Local Similarity   51.0%;  Pred. No. 13;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        251 AGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTA 310
              | ||||    | | ||| ||| | | | |||  | ||   |  |   ||   |  || |
Db      72532 AAACAGAGCCTAAAAACATTGGAAAGAATGAATAAGCCATCTTTTAAAAAATTAACAATT 72473

Qy        311 TTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAG 370
              ||   |||| |   |   ||  |      | ||||   ||  |   || |  ||| ||||
Db      72472 TTATCCAAGTTTTATTATGTTTTTAAATTAGGATTTATCAATATCCTTTTTTTAAAAAAG 72413

Qy        371 CAAATATTACACAGTATGGTACTCAAAAA 399
                . ||| | | || | ||  | | | ||| |
Db      72412 GAAACAATTCATAATAGTGGATTTAAACA 72384


RESULT 8
US-10-767-701-31198
; Sequence 31198, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 31198
;   LENGTH: 512
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(512)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 18065357
US-10-767-701-31198

  Query Match             6.9%;  Score 31.6;  DB 6;  Length 512;
  Best Local Similarity   50.7%;  Pred. No. 2.4;
  Matches   76;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        132 TCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGG 191
              | ||||   |||    | ||| | |||  | | | ||   |   ||     |||||||
Db        110 TGAGGCTCACATGGAAGTTCATGTTGGATCAATTGATGAGGAGGAGGACGAGCAGGGAAT 169

Qy        192 CTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCA 251
                 || |  |  | |    ||| |||      || ||||  || || | |||| ||
Db        170 TGCACATATGATTGAACATGTCGCGTTTCTTGGCAGTAAAAAACGTGAAAAGCTTTTGGG 229

Qy        252 GACAGGAGATTATAACCTTGCATATATTGA 281
              |||  || |          ||||||||  ||
Db        230 GACCGGTGCGAGGTCTAATGCATATACAGA 259


RESULT 9
US-10-417-884A-2325
; Sequence 2325, Application US/10417884A
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/417,884A
;             FILING DATE: 17-Apr-2003
;        PRIOR APPLICATION DATA:
```

```
;              APPLICATION NUMBER: US/09/107,532
;              FILING DATE: 30-Jun-1998
;              APPLICATION NUMBER: 60/085,598
;              FILING DATE: 14 May 1998
;              APPLICATION NUMBER: 60/051571
;              FILING DATE: July 2, 1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Ariniello, Pamela Deneke
;              REGISTRATION NUMBER: 40,489
;              REFERENCE/DOCKET NUMBER: GTC-012
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (781)893-5007
;              TELEFAX: (781)893-8277
;    INFORMATION FOR SEQ ID NO: 2325:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 525 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: double
;              TOPOLOGY: circular
;         MOLECULE TYPE: DNA (genomic)
;         HYPOTHETICAL: NO
;         ANTI-SENSE: NO
;         ORIGINAL SOURCE:
;              ORGANISM: Enterococcus faecium
;         FEATURE:
;              NAME/KEY:  misc_feature
;              LOCATION: (B) LOCATION 1...525
;         SEQUENCE DESCRIPTION: SEQ ID NO: 2325:
US-10-417-884A-2325

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 525;
  Best Local Similarity   54.9%;  Pred. No. 2.8;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        270 TGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGG 329
              || ||    |||| |||||     |    | ||||| ||||||   |||   |  ||
Db        213 TGAATTATCTGATAAGGCGAAAGAAGAAGCAGTTGCCATTATTTCTTCAGGGATTAGTGA 272

Qy        330 TAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACAC 382
                |||  ||  |||| |    | |    || |||  ||||||  || |||  ||
Db        273 CTATATTGATATGAGTGGTGAAATCAATTATGGCGATAAAGACAAAATTGTAC 325


RESULT 10
US-60-546-745-36
; Sequence 36, Application US/60546745
; GENERAL INFORMATION:
;  APPLICANT: Klaenhammer, Todd R.
;  APPLICANT:  Russell, William M.
;  APPLICANT:  Alterman, Eric
;  APPLICANT:  Cano, Raul
;  APPLICANT:  Hamrick, Alice
;  TITLE OF INVENTION: LACTOBACILLUS ACIDOPHILUS NUCLEIC ACID SEQUENCES ENCODING
;  TITLE OF INVENTION:  PROTEASE HOMOLOGUES AND USES THEREOF
;  FILE REFERENCE: 5051.604PR2
;  CURRENT APPLICATION NUMBER: US/60/546,745
;  CURRENT FILING DATE:  2004-02-23
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
;   LENGTH: 1245
;   TYPE: DNA
;   ORGANISM: Lactobacillus acidophilus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1245)
;   OTHER INFORMATION: pepT amino tripeptidase T (EC 3.411.-) ORF# 1190
US-60-546-745-36

  Query Match             6.9%;  Score 31.4;  DB 7;  Length 1245;
  Best Local Similarity   49.7%;  Pred. No. 3.7;
  Matches   80;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        240 AAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAA 299
              | |  ||||  |   |||      |  |  ||   || | | |  |||| || ||||  |
Db        129 AGATCTTGAGGAATTAGGCTTGAAAGATATTCACTACAATCAAAAGGCTGGTAGTGTAAT 188

Qy        300 CGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTC 359
              || ||  | |  |||  | | || |||||  |    |||  |||    ||  |  |  |
Db        189 CGCTGAAATTCCTTCAAATGTTGATTATGATGTACCTGTAATGGGATTCTTGGCACATAG 248

Qy        360 TGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAA 400
              || ||    |     | | || |  ||| ||  | | |||
Db        249 TGATACAGCTGATTTCAATTCAGAAAATGTTAAACCACAAA 289


RESULT 11
US-10-786-065-3
; Sequence 3, Application US/10786065
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001098DIV II
;  CURRENT APPLICATION NUMBER: US/10/786,065
;  CURRENT FILING DATE:  2004-02-26
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 53332
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-786-065-3

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 53332;
  Best Local Similarity   61.7%;  Pred. No. 13;
  Matches   50;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy        336 TGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCA 395
              ||  |||| ||| ||  |||  | ||   |||||||  ||| ||| |||  | |  |||
Db       8632 TGTTATGAGTATTCAACAAGAATATGTATATAAAGCACTTATCACAGAGTCAGTTTTTCA 8691

Qy        396 AAAAACGGCAATTGTAGTGCA 416
                |   |  ||| ||| |  ||
Db       8692 GAGTTCAACAAATGTTGACCA 8712


RESULT 12
US-10-767-471-8836/c
; Sequence 8836, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8836
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-8836

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 201;
  Best Local Similarity   54.6%;  Pred. No. 3.1;
  Matches   59;  Conservative    1;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGT 385
              |||| | ||  | |  ||  |||||  | |||  || ||||:| |    ||||  | |||
Db        142 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 83

Qy        386 ATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGG 433
              | | |  |||   |   |     | || ||||  ||||    |  |||
```

```
Db         82 AGGATTATCATGTACATGATGAGGCAGAGCAGGAACAGAAAAAGCTGG 35


RESULT 13
US-10-767-471-34493/c
; Sequence 34493, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34493
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-34493

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 201;
  Best Local Similarity   54.6%;  Pred. No. 3.1;
  Matches   59;  Conservative    1;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGT 385
              |||| | ||  | |  ||  |||||  | |||  || ||||:| |    |||| | |||
Db        142 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 83

Qy        386 ATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGG 433
              | | |  |||   |   |     | || ||||  ||||    |  |||
Db         82 AGGATTATCATGTACATGATGAGGCAGAGCAGGAACAGAAAAAGCTGG 35


RESULT 14
US-10-767-471-501/c
; Sequence 501, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 501
;   LENGTH: 1227
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-501

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 1227;
  Best Local Similarity   54.6%;  Pred. No. 5.7;
  Matches   59;  Conservative    1;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGT 385
              |||| | ||  | |  ||  |||||  | |||  || ||||:| |    |||| | |||
Db        476 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 417

Qy        386 ATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGG 433
              | | |  |||   |   |     | || ||||  ||||    |  |||
Db        416 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 369


RESULT 15
US-10-767-471-10718/c
; Sequence 10718, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10718
;   LENGTH: 76573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(76573)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10718

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 76573;
  Best Local Similarity   54.6%;  Pred. No. 22;
  Matches   59;  Conservative    1;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGT 385
              |||| | ||  | |  ||  |||||  | |||  || ||||:| |    |||| | |||
Db      69821 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 69762

Qy        386 ATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGG 433
              | | |  |||   |   |     | || ||||  ||||    |  |||
Db      69761 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 69714


Search completed: March 19, 2004, 03:10:03
Job time : 57.1445 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	205.8	45.1	220	12	BI941542	BI941542 dg26g06.y
c	2	56.8	12.5	207	9	AI368352	AI368352 SWOvMfCAR
c	3	47	10.3	680	12	BJ618688	BJ618688 BJ618688
c	4	40.2	8.8	956	29	CNS04QM8	AL302777 Tetraodon
	5	37.8	8.3	533	10	BB703456	BB703456 BB703456
	6	37	8.1	562	14	CF068694	CF068694 EST669415
	7	37	8.1	671	12	BI311207	BI311207 EST531295
	8	37	8.1	756	12	BM815787	BM815787 EST593881
c	9	36.6	8.0	957	13	BX453027	BX453027 BX453027
	10	36.4	8.0	364	12	BM540455	BM540455 qj59d01.b
	11	36.4	8.0	597	28	AZ169114	AZ169114 SP_0113_A
	12	36.4	8.0	657	10	BB449135	BB449135 BB449135
	13	36.4	8.0	769	12	BG354883	BG354883 MBTsMLA13
	14	36.2	7.9	491	28	AZ811111	AZ811111 2M0077A11
	15	36.2	7.9	532	28	AZ811109	AZ811109 2M0077A09
	16	36	7.9	678	12	BI452381	BI452381 GI09_G10
	17	36	7.9	818	9	AL716706	AL716706 AL716706
	18	35.6	7.8	615	28	AQ308392	AQ308392 CITBI-E1-
	19	35.6	7.8	624	28	AQ779365	AQ779365 HS_3207_B
	20	35.4	7.8	705	12	BG125943	BG125943 EST471589
	21	35.2	7.7	563	14	CA855727	CA855727 PfESToac3
c	22	35.2	7.7	744	12	BJ433259	BJ433259 BJ433259
c	23	35.2	7.7	1101	29	CNS0039G	AL063921 Drosophil
c	24	35.2	7.7	1134	9	AL551649	AL551649 AL551649
c	25	35	7.7	491	28	BH068121	BH068121 RPCI-24-3
c	26	34.8	7.6	774	28	BZ527786	BZ527786 OGAJU41TC
	27	34.8	7.6	788	14	CA766697	CA766697 AF53-Rpf_
	28	34.8	7.6	823	28	BZ527794	BZ527794 OGAJU41TM
	29	34.8	7.6	992	13	BU180244	BU180244 AGENCOURT
c	30	34.6	7.6	388	28	B86074	B86074 RPCI11-21O1
	31	34.6	7.6	572	13	BX552747	BX552747 BX552747
	32	34.6	7.6	1200	13	BX399167	BX399167 BX399167
	33	34.4	7.5	412	12	BM167028	BM167028 EST569551
c	34	34.4	7.5	413	9	AL698409	AL698409 DKFZp686I
	35	34.4	7.5	595	12	BM852853	BM852853 K-EST0134
c	36	34.4	7.5	849	29	CG950661	CG950661 MBEHO26TF
c	37	34.4	7.5	862	29	CNS029ER	AL187164 Tetraodon
c	38	34.4	7.5	876	28	BZ794472	BZ794472 PUFFR15TD
c	39	34.4	7.5	1050	13	BX349695	BX349695 BX349695
c	40	34.4	7.5	1201	13	BX324630	BX324630 BX324630
c	41	34.4	7.5	4397	29	AY413040	AY413040 Homo sapi
	42	34.2	7.5	528	10	BE599976	BE599976 PI1_78_A0
c	43	34.2	7.5	528	28	AQ606604	AQ606604 HS_2095_B
	44	34.2	7.5	531	12	BM323404	BM323404 PIC1_18_H
	45	34.2	7.5	623	29	CC816176	CC816176 1O0001A10

ALIGNMENTS

RESULT 1
BI941542
LOCUS BI941542 220 bp mRNA linear EST 19-FEB-2003
DEFINITION dg26g06.y1 Xenopus laevis gastrula non normalized Xenopus laevis cDNA clone IMAGE:3749963 5' similar to SW:CSGB_ECOLI P39828 MINOR CURLIN SUBUNIT PRECURSOR. ;, mRNA sequence.
ACCESSION BI941542
VERSION BI941542.1 GI:16256014
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 220)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Source lab clone id - 3749963
Trace considered overall poor quality This clone is available royalty-free through LLNL; contact the IMAGE Consortium (image@image.llnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .220
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:3749963"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F'"
/clone_lib="Xenopus laevis gastrula non normalized"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2 ug of poly A+ RNA (equal parts from stage 10.5 and stage 11.5 gastrulae). EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. The library was mass excised and used to infect Top10F'. Clones were picked into freezing medium (per liter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM K2HPO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4 7 H2O, 6.8 mM (NH4)2SO4, 4 % w/v glycerol) and grown for 24 hours. Original library construction by Bruce Blumberg (Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

Query Match 45.1%; Score 205.8; DB 12; Length 220;
Best Local Similarity 96.8%; Pred. No. 4.7e-51;
Matches 210; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy         80 CTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAG 139
               ||||||| |||||||||||||||| | |||||| |||||||||||||||||||||||||
Db          3 CTAATTCATAATATAACTTCGCGGCACATGAATCGAGTAAGTCTTCATTTAATCAGGCAG 62

Qy        140 CCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAAC 199
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         63 CCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACTGCAGGGAGGCTCAAAAC 122

Qy        200 TTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAG 259
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TTTTGGCGGTTGCTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAG 182

Qy        260 ATTATAACCTTGCATATATTGATCAGGCGGGCAGTGC 296
              |||||||||||||||||||||||||||||||||||| |
Db        183 ATTATAACCTTGCATATATTGATCAGGCGGGCAGTCC 219
```

RESULT 2
AI368352/c
LOCUS AI368352 207 bp mRNA linear EST 11-JAN-1999
DEFINITION SWOvMfCAR04G09SK Onchocerca volvulus microfilaria cDNA (SAW98MLW-OvMf) Onchocerca volvulus cDNA clone SWOvMfCAR04G09 5', mRNA sequence.
ACCESSION AI368352
VERSION AI368352.1 GI:4147105
KEYWORDS EST.
SOURCE Onchocerca volvulus
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 207)
AUTHORS Williams,S.A.
TITLE Genes expressed in microfilaria of Onchocerca volvulus
JOURNAL Unpublished (1999)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.

FEATURES Location/Qualifiers
source 1. .207
/organism="Onchocerca volvulus"
/mol_type="mRNA"
/db_xref="taxon:6282"
/clone="SWOvMfCAR04G09"
/dev_stage="microfilaria"
/lab_host="XL1-Blue MRF'"
/clone_lib="Onchocerca volvulus microfilaria cDNA (SAW98MLW-OvMf)"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.8 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email:genome@smith.edu."

ORIGIN

Query Match 12.5%; Score 56.8; DB 9; Length 207;
Best Local Similarity 95.1%; Pred. No. 3.9e-06;
Matches 58; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||| | |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGAAAACCNAATTGTCATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 2

Qy         61 G 61
              |
Db          1 G 1
```

RESULT 3
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.
http://xenopus.nibb.ac.jp.

FEATURES Location/Qualifiers
source 1. .680

```
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match             10.3%;  Score 47;  DB 12;  Length 680;
  Best Local Similarity   100.0%;  Pred. No. 0.005;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        410 TAGTGCAGAGACAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        680 TAGTGCAGAGACAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 634


RESULT 4
CNS04QM8/c
LOCUS       CNS04QM8                 956 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            129D18 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL302777
VERSION     AL302777.1  GI:8183119
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 956)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="129D18"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG129DB09SP1-end :
                     PUC-Ori"
ORIGIN
```

```
  Query Match             8.8%;  Score 40.2;  DB 29;  Length 956;
  Best Local Similarity   46.6%;  Pred. No. 0.62;
  Matches  129;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        113 TGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTG 172
              | ||| ||||   |   | | |   | |  | ||  ||||  | |||   | ||||
Db        430 TCAGTCAGTCAGGAGCCACTAATCAATCACTGATCAGTCAGTCAGGGGTCACTAATCAAT 371

Qy        173 CTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCA 232
              | | | |  | |||  |||  | |   |   |  |  |   ||   |||    | |
Db        370 CACTGATCAGTCAGTCAGGAGCCACTAATCAATCACTGATCAGTCAGTCAGGGGTCACTA 311

Qy        233 ACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCA 292
              | |   ||  |||    ||| ||||||   ||| |   ||   ||   |||| | ||
Db        310 ATCAATCACTGATCAGTCAGTCAGGAGTCACTAATCAATCACTGATCAGTCAGTCAGGGG 251

Qy        293 GTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGA 352
              | ||  | |    ||    ||    |       ||||    | |||| |  |||
Db        250 TCACTAATCAATCACTGATCAGTCAGTCAGGGGTCACTAATCAATCACTGATCAGTCAGT 191

Qy        353 AAGGTTCTGGTAATAAAGCAAATATTACACAGTATGG 389
              |||  |  |||| || ||  || |  ||||  ||
Db        190 CAGGAGCCACTAATCAATCACTGATCAGTCAGTCAGG 154


RESULT 5
BB703456
LOCUS       BB703456                 533 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  BB703456 RIKEN full-length enriched, in vitro fertilized eggs Mus
            musculus cDNA clone 7420452F19 3', mRNA sequence.
ACCESSION   BB703456
VERSION     BB703456.1  GI:16052291
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
            Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
```

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .533
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="7420452F19"
/sex="female"
/tissue_type="in vitro fertilized eggs"
/dev_stage="egg"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, in vitro fertilized eggs"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match 8.3%; Score 37.8; DB 10; Length 533;
Best Local Similarity 49.3%; Pred. No. 2.8;
Matches 99; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
Qy      126 ATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCA 185
            ||||  | || | | ||||||    || |||  || |  ||| | |     || |  |
Db      202 ATTTTTTTAGCCGGACATAATGTTGCATGCTTTGAATCCTAACATTCAGAAGCCAGAGGC 261

Qy      186 GGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGAT 245
              || ||      |  | |    ||  ||| | ||  | |     |      || ||| || | |
Db      262 AGGTGGAGATCTATATATACAGGGAGTTCCTCACTAGGCAGGGCTACACTGGGAAATGCT 321

Qy      246 TGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGC 305
              ||   ||  | ||| |    ||| || |||||||  |  ||    ||   || | ||
Db      322 GCCCCCCCCAAAAAATTGTTGAGTTGTATTTATTGATTTGAGGGCTTATGTAGACCAGGC 381

Qy      306 CAGTATTTCGCAAGGTGCTTA 326
            |   ||    | | ||  ||
Db      382 TATCCTTGAACTTGTTGTCTA 402
```

RESULT 6
CF068694
LOCUS CF068694 562 bp mRNA linear EST 21-JUL-2003
DEFINITION EST669415 MTUS Medicago truncatula cDNA clone MTUS-11C11, mRNA sequence.
ACCESSION CF068694
VERSION CF068694.1 GI:33105353
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 562)
AUTHORS VandenBosch,K., Endre,G., Silverstein,K, Town,C.D., Van Aken,S., Utterback,T., Cheung,F. and Fraser,C.M.
TITLE The Medicago truncatula 6K unigene set: cDNA clones selected and re-arrayed from various libraries
JOURNAL Unpublished (2002)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
TIGR sequence name: MTUSK35TK Alias Clone name:pGESD10A6 More information is available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .562
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="MTUS-11C11"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/clone_lib="MTUS"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 562;
Best Local Similarity 53.9%; Pred. No. 4.9;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy      242 AGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACG 301
            |||||||  ||||  |  |||  || |  | |  || |  |||||      | ||    |
Db       59 AGATTGAACAGAGTGATGATAATGATGTGGATTACAAAAATCAGCTTCTTAATGTTCTCA 118

Qy      302 ATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTG 361
            |   |||  |    | |||  || |  |  || |||  ||||||  ||    |  | | | |
Db      119 AAATCAGCAAGGAGAAAGTTGATGAATGTTATAATGCGATTCTTCATCTTACAAATACAA 178

Qy      362 GTAATAAAGCAAATATTACAC 382
            ||||  | |   ||| || ||
Db      179 ATAATTATGGTTATAATAAAC 199
```

RESULT 7
BI311207
LOCUS BI311207 671 bp mRNA linear EST 20-JUL-2001
DEFINITION EST5312957 GESD Medicago truncatula cDNA clone pGESD10A6 5' end, mRNA sequence.
ACCESSION BI311207
VERSION BI311207.1 GI:14985534
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 671)
AUTHORS Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from developing reproductive tissues of Medicago truncatula
JOURNAL Unpublished (2001)
COMMENT Contact: Michael A. Grusak
USDA/ARS Children's Nutrition Research Center

Baylor College of Medicine
1100 Bates Street, Houston, TX 77030-2600, USA
Tel: 713-798-7044
Fax: 713-798-7078
Email: mgrusak@bcm.tmc.edu
B398062e
TIGR sequence name: MTPAR03TK
More information is available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .671
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pGESD10A6"
/tissue_type="immature seeds"
/dev_stage="Immature seeds, 11 to 19 days after pollination"
/clone_lib="GESD"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Immature seeds, collected from pods ranging in age from 11 to 19 days after pollination, were harvested from greenhouse-grown plants. Seeds were removed and separated from pod walls and were immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propogated in XLOLR cells."

ORIGIN

Query Match 8.1%; Score 37; DB 12; Length 671;
Best Local Similarity 53.9%; Pred. No. 5.1;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        242 AGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACG 301
              ||||||| ||||  |  ||| || |  | |  || |   |||||      | ||    |
Db         79 AGATTGAACAGAGTGATGATAATGATGTGGATTACAAAAATCAGCTTCTTAATGTTCTCA 138

Qy        302 ATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTG 361
              |   ||| |    | ||| || | |  || ||| ||||||  ||   |  | |  | |
Db        139 AAATCAGCAAGGAGAAAGTTGATGAATGTTATAATGCGATTCTTCATCTTACAAATACAA 198

Qy        362 GTAATAAAGCAAATATTACAC 382
               |||| | |   ||| || ||
Db        199 ATAATTATGGTTATAATAAAC 219
```

RESULT 8
BM815787
LOCUS BM815787 756 bp mRNA linear EST 05-MAR-2002
DEFINITION EST593881 BNIR Medicago truncatula/Meloidogyne incognita mixed EST library cDNA clone pBNIR-28N9, mRNA sequence.
ACCESSION BM815787
VERSION BM815787.1 GI:19151801
KEYWORDS EST.
SOURCE Medicago truncatula/Meloidogyne incognita mixed EST library
ORGANISM Medicago truncatula/Meloidogyne incognita mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 756)
AUTHORS Bird,D., Koltai,H., Samac,D., Town,C.D., Van Aken,S., Utterback,T., Cheung,F., Tsai,J. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula after infection with the nematode Meloidogyne incognita
JOURNAL Unpublished (2002)
COMMENT Contact: Bird, DM
Plant Nematode Genetics Group
North Carolina State University
Box 7616, Raleigh, NC 27695, USA
Tel: 919 515 6813
Fax: 919 515 9500
Email: david_bird@ncsu.edu
TIGR sequence name: MTQBI77TK More information is available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .756
/organism="Medicago truncatula/Meloidogyne incognita mixed EST library"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:188702"
/clone="pBNIR-28N9"
/tissue_type="3 week old roots"
/dev_stage="3 days after infection with Meloidogyne incognita second stage larvae"
/lab_host="XLOLR"
/clone_lib="BNIR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells. "

ORIGIN

Query Match 8.1%; Score 37; DB 12; Length 756;
Best Local Similarity 53.9%; Pred. No. 5.3;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        242 AGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACG 301
              ||||||| ||||  |  ||| || |  | |  || |   |||||      | ||    |
Db        251 AGATTGAACAGAGTGATGATAATGATGTGGATTACAAAAATCAGCTTCTTAATGTTCTCA 310

Qy        302 ATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTG 361
              |   ||| |    | ||| || | |  || ||| ||||||  ||   |  | |  | |
Db        311 AAATCAGCAAGGAGAAAGTTGATGAATGTTATAATGCGATTCTTCATCTTACAAATACAA 370

Qy        362 GTAATAAAGCAAATATTACAC 382
               |||| | |   ||| || ||
Db        371 ATAATTATGGTTATAATAAAC 391
```

RESULT 9
BX453027/c
LOCUS BX453027 957 bp mRNA linear EST 22-MAY-2003
DEFINITION BX453027 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF012YJ13 3-PRIME, mRNA sequence.
ACCESSION BX453027
VERSION BX453027.1 GI:31036338
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
was not normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6669.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAH007ZA11NP1&cluster=6669.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0BAH007ZA11NP1.

```
FEATURES             Location/Qualifiers
     source          1. .957
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF012YJ13"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             8.0%;  Score 36.6;  DB 13;  Length 957;
  Best Local Similarity   32.1%;  Pred. No. 7.5;
  Matches  106;  Conservative   55;  Mismatches  169;  Indels    0;  Gaps    0;

Qy         98 TCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTG 157
Db        920 TGGGGGAGANGGGARWKATTAAKAVKGGGGDAAAKGAKKWTAAAAAAAKGGAGAAGGTKG 861

Qy        158 GGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGC 217
Db        860 GGTTATTKGASSSSGGRSAAAAASGGKAWKAKTKAGAAAAWKAGGTGGTGAAAAATKKTA 801

Qy        218 AAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATA 277
Db        800 KTAWKGGGKGKRAAAAWKAASAAKTDGGAMAAAARATKAAAGTTATKAAGWATGVAAKAA 741

Qy        278 TTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTG 337
Db        740 KGAAAAAAAKAAKGWKKAGAAAAAAKGGAAAAGGKAAKAAAKAAAAAAKGGGGKAKAGKG 681

Qy        338 CGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAA 397
Db        680 GAAGAAWKAATGAAAAAAGKAAAKGAAWTAAAKGAAAAAAAGAAGAAKGKGGKKGGAAAA 621

Qy        398 AAACGGCAATTGTAGTGCAGAGACAGTCGC 427
Db        620 TAASGGAWATKGKGGAGKSAKGGAASASSS 591


RESULT 10
BM540455
LOCUS       BM540455                 364 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  qj59d01.b1 Canis cDNAs from testes cells Canis familiaris cDNA
            clone qj59d01 5', mRNA sequence.
ACCESSION   BM540455
VERSION     BM540455.1  GI:18823255
KEYWORDS    EST.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   O'Shaughnessy,A.L., McCombie,W.R., Baker,J.P., Balija,V.,
            Cunnius,D., Dedhia,N.N., de la Bastide,M., Katzenberger,F.,
            King,L., Kirchoff,K.A., Miller,B., Muller,S., Nascimento,L.U.,
            Palmer,L., Santos,L., Shah,R.S., Spiegel,L.A., Zutavern,T.,
            Preston,R. and Hannon,G.J.
  TITLE     Expressed sequence tags from Canis familiaris (dog) (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: qj59  row: d  column: 01
            Seq primer: -21M13UnivFwd
            High quality sequence stop: 364.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Canis familiaris"
                     /mol_type="mRNA"
                     /db_xref="taxon:9615"
                     /clone="qj59d01"
                     /tissue_type="testes"
                     /clone_lib="Canis cDNAs from testes cells"
                     /note="Vector: Lambda Zap II; The library was produced by
                     Greg Hannon and Raymond Preston (Cold Spring  Harbor
                     Laboratory). This library is oligo(dT) primed using
                     stratagene  zap cDNA synthesis kit. It was made from dog
                     testes. Please contact Greg  Hannon (hannon@cshl.org) with
                     any library related inquiries."
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 12;  Length 364;
  Best Local Similarity   49.0%;  Pred. No. 6.5;
  Matches   97;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        231 CAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGG 290
Db        101 CCAGGGGGGAACCTTTGAACAAGAGGGTCTCCAGCATCTTTGAAATGCTCAGTTGATGGG 160

Qy        291 CAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCA 350
Db        161 GAAAACAAACTTTGTAGTCATTTGGCTAGGGAAATTGGCTTTAACCCCTCCTCCTACTCT 220

Qy        351 GAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGT 410
Db        221 TTGTTCCTATCCTAAAAGACTAACCATTTACCCCTACTCTACTGAAAAAGGGGCTGGTGT 280

Qy        411 AGTGCAGAGACAGTCGCA 428
Db        281 AGGGCCACGACAGTCTCA 298


RESULT 11
AZ169114
LOCUS       AZ169114                 597 bp    DNA     linear   GSS 29-AUG-2000
DEFINITION  SP_0113_A2_A09_T7A Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library Strongylocentrotus purpuratus
            genomic clone Plate=113 Col=18 Row=A, genomic survey sequence.
ACCESSION   AZ169114
VERSION     AZ169114.1  GI:8339482
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
            Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H.
            and Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
   PUBMED   10920195
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 113  row: A  column: 18
            Seq primer: T7
            Class: BAC ends
```

```
          High quality sequence stop: 597.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7668"
                     /clone="Plate:113 Col:18 Row:A"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 28;  Length 597;
  Best Local Similarity   51.2%;  Pred. No. 7.5;
  Matches   85;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        129 TAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGG 188
              || |||     || |  ||||||| | |||   || | || || |     |  | |||
Db        233 TAGTCATCGTCCCTTTCTTGGTCACGATGGCCATAGTCATTGTCCCTTTCTTGGTCAGCC 292

Qy        189 AGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGA 248
              | ||  |  | |     |  || |||  ||     ||| ||||  || || ||      ||
Db        293 TGTCTGTAGTCATCGTCCCTTTCTTGGTCAGCCTGGTCGTAGTCACTGGCCATTTCCTGC 352

Qy        249 CCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGT 294
              ||  | |     |  | | | | | | |  || |||| | ||| ||
Db        353 TCATCCTGTCTGTAGTCATCGTCCCTTTCCTGGTCAGCCTGGCCGT 398


RESULT 12
BB449135
LOCUS       BB449135                657 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  BB449135 RIKEN full-length enriched, 9 days embryo Mus musculus
            cDNA clone D030070B09 3' similar to D38046 Mouse mRNA for typeII
            DNA topoisomerase beta isoform, mRNA sequence.
ACCESSION   BB449135
VERSION     BB449135.2  GI:16425330
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 21, 2000 this sequence version replaced gi:9310170.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="D030070B09"
                     /dev_stage="9 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 9 days embryo"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 10;  Length 657;
  Best Local Similarity   50.6%;  Pred. No. 7.7;
  Matches   88;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
              || ||  || ||| |||  |     || | | |||  |  |   | || | |   | ||
Db        478 AGAAGATACAGGGGAAAATTACTATAGGCTGCAGAAGAGGAAGATTCACAAAACCAACAT 537

Qy        286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
              |  | |||| ||   ||| |  | | | |   |||  || ||  |||| |    | | |
Db        538 GATGACAGTTCCTCTGATTCAGGAACTCCTTCAGGCCCTGATTTTAATTATATTTTAAAT 597

Qy        346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAA 399
              ||       | | |||  || ||||| |||  ||  | |       ||  || | |
Db        598 ATGTCTCTATGGTCTCTTACTAAAGAAAAAGTTGAAGAACTCATTAAACAGAGA 651


RESULT 13
BG354883
LOCUS       BG354883                769 bp    mRNA    linear   EST 05-MAR-2001
DEFINITION  MBTsMLA133T7SEQ Trichinella spiralis muscle stage larvae (BC)
            Trichinella spiralis cDNA clone MBTsMLA133 5', mRNA sequence.
ACCESSION   BG354883
VERSION     BG354883.1  GI:13199069
KEYWORDS    EST.
SOURCE      Trichinella spiralis
  ORGANISM  Trichinella spiralis
            Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida;
```

```
            Trichinellidae; Trichinella.
REFERENCE   1  (bases 1 to 769)
  AUTHORS   Daub,J., Connolly,B., Garate,T. and Blaxter,M.L.
  TITLE     A survey of genes expressed in the muscle stage larvae of the
            parasitic nematode Trichinella spiralis
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh,
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK.
            Tel: +44 131 650 6760
            Fax: +44 131 670 5450
            Email: mark.blaxter@ed.ac.uk
            PCR PRimers
            FORWARD: T7SEQ (GTAATACGACTCACTATAGGG)
            BACKWARD: M13 Forward (CGCCAGGGTTTTCCCAGTCACGAC)
            Seq primer: T7SEQ (GTAATACGACTCACTATAGGG).
FEATURES                Location/Qualifiers
     source             1. .769
                        /organism="Trichinella spiralis"
                        /mol_type="mRNA"
                        /db_xref="taxon:6334"
                        /clone="MBTsMLA133"
                        /sex="mixed"
                        /dev_stage="muscle stage larvae"
                        /clone_lib="Trichinella spiralis muscle stage larvae (BC)"
                        /note="Vector: Lambda ZipLox; Site_1: SalI (5'end);
                        Site_2: NotI (3'end); The infective L1 larva of
                        Trichinella spiralis is a nematode parasite of mammalian
                        skeletal muscle. The library was constructed using muscle
                        larvae of the T. spiralis isolate ISS3, and was provided
                        by Dr Bernadette Connolly, University of Aberdeen."
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 12;  Length 769;
  Best Local Similarity   50.6%;  Pred. No. 8.1;
  Matches   88;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        239 CAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCA 298
              ||   ||  | | || |  ||| |||  |  || || |    |    |    | | | |
Db         78 CAGTTATGAATCGGATAACAGAAAATATACGCGCTTACAAATTTGGTCCATCGACTTCAA 137

Qy        299 ACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTT 358
              |||||| || |||    ||| || ||    ||  |||        ||        | ||
Db        138 TCGATGCAAGCATTGATGAAGATGATTTACCTATGACTTATGGATTTGAAGCTTTGGTTT 197

Qy        359 CTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAG 412
              ||| | ||||||  ||||||    |  |     |  ||||| | | ||||| ||
Db        198 CTGATGATAAAGGCAATATTGTTAATGAACTGATGAAAAAATCAGTAATTGGAG 251


RESULT 14
AZ811111
LOCUS       AZ811111                491 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0077A11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0077A11 F, genomic survey sequence.
ACCESSION   AZ811111
VERSION     AZ811111.1  GI:12979042
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0077  row: A  column: 11
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 491.
FEATURES                Location/Qualifiers
     source             1. .491
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="UUGC2M0077A11"
                        /sex="Male"
                        /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                        /clone_lib="Mouse 10kb plasmid UUGC1M library"
                        /note="Vector: PWD42nv; Purified genomic DNA from M.
                        musculus C57BL/6J (male) was obtained  from the Jackson
                        Laboratory Mouse DNA Resource
                        (http://www.jax.org/resources/documents/dnares/). The DNA
                        was hydrodynamically sheared by repeated passage through a
                        0.005 inch orifice at constant velocity. The sheared DNA
                        was blunt end-repaired with T4 DNA polymerase and T4
                        polynucleotide kinase. Adaptor oligonucleotides were
                        ligated to the blunt ends in high  molar excess. The
                        adaptored DNA was purified and size-selected for a 9.5 to
                        10.5 kb range using preparative agarose gel
                        electrophoresis. Vector DNA was prepared from a derivative
                        of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                        inducible derivative of plasmid R1. The vector was ligated
                        with adaptors complementary to the insert adaptors and
                        purified. The sheared, adaptored mouse DNA was annealed to
                        adaptored vector DNA, and transformed into
                        chemically-competent E. coli XL10-Gold (Stratagene) cells
                        and selected for ampicillin resistance."
ORIGIN

  Query Match             7.9%;  Score 36.2;  DB 28;  Length 491;
  Best Local Similarity   52.3%;  Pred. No. 8.2;
  Matches   80;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        270 TGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGG 329
              ||   |||   ||      | || ||| || | | | |||||| |||     || |||  |
Db        110 TGAGGATAAATATTTTCAGAGCTGTGTCAGCAAAGGCAGTATATCGTTTCATGTTTAGAG 169

Qy        330 TAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGG 389
              ||     |||||  || | | | | |   |||||   ||| ||  |  | | | |  |
Db        170 TACATGTGCGATAGTTCTGGAAATATGGATTGGTATACAAGAAAGAAAAAGAAAAAAGAA 229

Qy        390 TACTCAAAAAACGGCAATTGTAGTGCAGAGACA 422
              |  |||||  |  ||  |  |  | | ||| |
Db        230 AAAAAGAAAAAAGAGAAAAGAATAGAAAAGAAA 262


RESULT 15
AZ811109
LOCUS       AZ811109                532 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0077A09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0077A09 F, genomic survey sequence.
ACCESSION   AZ811109
VERSION     AZ811109.1  GI:12979038
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 532)
```

```
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0077  row: A  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 532.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0077A09"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

  Query Match             7.9%;  Score 36.2;  DB 28;  Length 532;
  Best Local Similarity   52.3%;  Pred. No. 8.4;
  Matches   80;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        270 TGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGG 329
              ||   |||   ||     | || ||| || | | | |||||| |||     || |||  |
Db         92 TGAGGATAAATATTTTCAGAGCTGTGTCAGCAAAGGCAGTATATCGTTTCATGTTTAGAG 151

Qy        330 TAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGG 389
              ||     ||||||  || |  | | | |   |||||   ||| ||  |  | | |  |
Db        152 TACATGTGCGATAGTTCTGGAAATATGGATTGGTATACAAGAAAGAAAAAGAAAAAAGAA 211

Qy        390 TACTCAAAAAACGGCAATTGTAGTGCAGAGACA 422
               |    ||||| |  ||  | |  | | ||| |
Db        212 AAAAAGAAAAAAGAGAAAAGAATAGAAAAGAAA 244


Search completed: March 18, 2004, 20:11:20
Job time : 2038.55 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 243.087 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	29.4	61.2	1797	3	LEIMSP52A	L19563 Leishmania
2	29.4	61.2	1860	3	LIGP63GEN	Z83677 L.infantum
3	29.4	61.2	2104	3	LDO495002	AJ495002 Leishmani
4	29.4	61.2	2106	3	LDO495006	AJ495006 Leishmani
5	29.4	61.2	2106	3	LDO495007	AJ495007 Leishmani
6	29.4	61.2	2107	3	LDO495005	AJ495005 Leishmani
7	29.4	61.2	2109	3	LDO495003	AJ495003 Leishmani
8	29.4	61.2	2109	3	LDO495004	AJ495004 Leishmani
9	29.4	61.2	2892	3	LEIGP63D	M80672 Leishmania
10	29.4	61.2	2966	3	LEIGP63B	M80669 Leishhmania
11	29.4	61.2	3047	3	LEIGP63A	M28527 L.chagasi m
12	29.4	61.2	3047	3	LIU48798	U48798 Leishmania
13	29.4	61.2	3105	3	LEIGPAA	M60048 L.donovani.
14	28.4	59.2	2621	3	LEIGP63C	M80671 Leishmania
15	27.8	57.9	2050	3	LTU495010	AJ495010 Leishmani
16	27.8	57.9	2058	3	LAE532831	AJ532831 Leishmani
17	27.8	57.9	2077	3	LTR495009	AJ495009 Leishmani
18	27.8	57.9	2090	3	LAR532830	AJ532830 Leishmani
19	27.8	57.9	2125	3	LTR495008	AJ495008 Leishmani
20	27.8	57.9	2161	3	LMGP63	Y00647 Leishmania
21	27.8	57.9	2187	3	LMGP63C1	X64394 L.mexicana
22	27.8	57.9	3229	3	AF039721	AF039721 Leishmani
c 23	27.8	57.9	11523	1	AE014950	AE014950 Streptoco
c 24	26.8	55.8	57112	2	AC019976	AC019976 Drosophil
c 25	26.8	55.8	88031	2	AC004378	AC004378 Drosophil
26	26.8	55.8	124700	3	AC005558	AC005558 Drosophil
27	26.8	55.8	126807	9	HS391022	AL031577 Human DNA
c 28	26.8	55.8	171774	3	AC009909	AC009909 Drosophil
c 29	26.8	55.8	295134	3	AE003582	AE003582 Drosophil
30	26.4	55.0	1800	3	LEIMSPS4A	L19562 Leishmania
31	26.2	54.6	2046	3	LEIGP63E	L46798 Leishmania
c 32	25.8	53.8	238365	2	AC107269	AC107269 Rattus no
33	24.6	51.3	1964	3	LIGP63	Y08156 L.infantum
34	24.6	51.3	2544	6	AX568314	AX568314 Sequence
c 35	24.6	51.3	3636	6	AR218846	AR218846 Sequence
c 36	24.6	51.3	3636	6	BD003758	BD003758 Polynucle
37	24.6	51.3	11495	1	AE007386	AE007386 Streptoco
38	24.6	51.3	12092	1	AE008447	AE008447 Streptoco
39	24.6	51.3	147006	2	SPNEU1904	AL449926 Streptoco
40	24.6	51.3	349980	6	AX571761	AX571761 Sequence
41	24.6	51.3	349980	6	AX571762	AX571762 Sequence
c 42	24.4	50.8	2240	8	AY118173	AY118173 Ustilago
43	24.2	50.4	144239	10	AL627347	AL627347 Mouse DNA
c 44	24.2	50.4	182340	2	AC099639	AC099639 Mus muscu
c 45	24.2	50.4	223974	2	AC106343	AC106343 Rattus no

ALIGNMENTS

RESULT 1
LEIMSP52A
LOCUS LEIMSP52A 1797 bp DNA linear INV 28-NOV-1994
DEFINITION Leishmania donovani glycoprotein 63 (mspS2) gene, complete cds.
ACCESSION L19563
VERSION L19563.1 GI:308887
KEYWORDS glycoprotein 63; surface protein; virulence factor; zinc protease.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 1797)
AUTHORS Roberts,S.C., Swihart,K.G., Agey,M.W., Ramamoorthy,R., Wilson,M.E. and Donelson,J.E.
TITLE Sequence diversity and organization of the msp gene family encoding

gp63 of Leishmania chagasi
JOURNAL Mol. Biochem. Parasitol. 62 (2), 157-171 (1993)
MEDLINE 94187791
PUBMED 8139613
COMMENT Original source text: Leishmania donovani (sub_species chagasi) stationary stage promastigote DNA.

FEATURES Location/Qualifiers
source 1. .1797
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
/dev_stage="stationary stage promastigote"
gene 1. .1797
/gene="mspS2"
CDS 1. .1797
/gene="mspS2"
/note="propeptide"
/codon_start=1
/product="glycoprotein 63"
/protein_id="AAA53688.1"
/db_xref="GI:308888"
/translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEG
DVLAWAATCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSDTFFTD
KRMLHNVGKIRGKPHNAPVINSSTAVAKAREQYGCDTLEYLEMEDEGGAGSAGSHIKM
RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
DGITKWPAMFCNSDDALRCPTSRLSLGACSLATYQSLPPYWQYFTDPSLAGISAFMDY
CPVVVPFGNGSCAQNASKVMAAVQAFNVFSDAARCIDGAFRPKTTETVTNSYAGLCAN
VLCDTATRTYSVQVRGGSGYASCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNVQ
AAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
sig_peptide 1. .39
/gene="mspS2"
mat_peptide 97. .1794
/gene="mspS2"
/product="glycoprotein 63"

ORIGIN

Query Match 61.2%; Score 29.4; DB 3; Length 1797;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
|| || |||||||| || ||||| || || || || ||||| || ||
Db 751 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 797

RESULT 2
LIGP63GEN
LOCUS LIGP63GEN 1860 bp DNA linear INV 16-DEC-1996
DEFINITION L.infantum GP63 gene.
ACCESSION Z83677
VERSION Z83677.1 GI:1743284
KEYWORDS gp63 gene; protease.
SOURCE Leishmania infantum
ORGANISM Leishmania infantum
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
REFERENCE 1 (bases 1 to 1860)
AUTHORS Ortiz,G., Martinez,P. and Segovia,M.
TITLE Characterization of GP63 of Leishmania infantum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1860)
AUTHORS Ortiz,G.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1996) Ortiz G., Facultad de Medicina. Universidad de Murcia, Genetica y Microbiologia, Campus de Espinardo, Murcia, Murcia, SPAIN, 30100

FEATURES Location/Qualifiers
source 1. .1860
/organism="Leishmania infantum"
/mol_type="genomic DNA"
/strain="cutaneous"
/db_xref="taxon:5671"
/dev_stage="promastigote"
CDS 16. .1815
/function="protease"
/codon_start=1
/product="GP63"
/protein_id="CAB06018.1"
/db_xref="GI:1743285"
/translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGIIKSYAGLCA
NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"

ORIGIN

Query Match 61.2%; Score 29.4; DB 3; Length 1860;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
|| || |||||||| || ||||| || || || || ||||| || ||
Db 766 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 812

RESULT 3
LDO495002
LOCUS LDO495002 2104 bp DNA linear INV 22-JUL-2002
DEFINITION Leishmania donovani mspC gene for GP63, strain MHOM/KE/0000/Neal-R1.
ACCESSION AJ495002
VERSION AJ495002.1 GI:21954455
KEYWORDS GP63; major surface protease; mspC gene.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
REFERENCE 1
AUTHORS Mauricio,I.L., Stothard,J.R. and Miles,M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2104)
AUTHORS Mauricio,I.L.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM

FEATURES Location/Qualifiers
source 1. .2104
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="MHOM/KE/0000/Neal-R1"
/db_xref="taxon:5661"
/country="Kenya"
gene 15. .1940
/gene="mspC"
CDS 15. .1940
/gene="mspC"
/note="constitutive major surface protease"
/codon_start=1
/allele="mspCLdA1"
/product="GP63"
/protein_id="CAD42811.1"
/db_xref="GI:21954456"
/db_xref="GOA:Q8MNZ5"
/db_xref="SPTREMBL:Q8MNZ5"

```
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVACHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDSAYHCARVGQRVNNHAGAIATCTAEDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWNVTDMVDEICGDFKVPQAHITEGFSNTDFVMYVASVPSEK
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILLVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVRGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQAN
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     15. .131
                     /gene="mspC"
     mat_peptide     132. .1823
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match            61.2%;  Score 29.4;  DB 3;  Length 2104;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        765 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 811


RESULT 4
LDO495006
LOCUS       LDO495006               2106 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain MCAN/IQ/1977/Bumm3.
ACCESSION   AJ495006
VERSION     AJ495006.1  GI:21954463
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2106)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MCAN/IQ/1977/Bumm3"
                     /db_xref="taxon:5661"
                     /country="Iraq"
     gene            14. .1939
                     /gene="mspC"
     CDS             14. .1939
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdG2"
                     /product="GP63"
                     /protein_id="CAD42815.1"
                     /db_xref="GI:21954464"
                     /db_xref="GOA:Q8MM48"
                     /db_xref="SPTREMBL:Q8MM48"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMHPRVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTADDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWKVTDMVDEICGDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILVVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCGTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVCGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQAN
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     14. .130
                     /gene="mspC"
     mat_peptide     131. .1822
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match            61.2%;  Score 29.4;  DB 3;  Length 2106;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        764 TACGACCAGCTGGTGACACGTGTCGTCACGCACGAGATGGCGCACGC 810


RESULT 5
LDO495007
LOCUS       LDO495007               2106 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain
            MHOM/SU/1984/Marz-Krim.
ACCESSION   AJ495007
VERSION     AJ495007.1  GI:21954465
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2106)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MHOM/SU/1984/Marz-Krim"
                     /db_xref="taxon:5661"
     gene            14. .1939
                     /gene="mspC"
     CDS             14. .1939
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdH"
                     /product="GP63"
                     /protein_id="CAD42816.1"
                     /db_xref="GI:21954466"
                     /db_xref="GOA:Q8MM48"
                     /db_xref="SPTREMBL:Q8MM48"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMHPRVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTADDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWKVTDMVDEICGDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILVVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCGTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVCGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQAN
```

```
                VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                LLLTCPWWCCKFGGLPT"
     sig_peptide 14. .130
                /gene="mspC"
     mat_peptide 131. .1822
                /gene="mspC"
                /product="GP63"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 2106;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        764 TACGACCAGCTGGTGACACGTGTCGTCACGCACGAGATGGCGCACGC 810


RESULT 6
LDO495005
LOCUS       LDO495005               2107 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain
            MCAN/IQ/1981/Sukkar2.
ACCESSION   AJ495005
VERSION     AJ495005.1  GI:21954461
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2107)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2107
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MCAN/IQ/1981/Sukkar2"
                     /db_xref="taxon:5661"
                     /country="Iraq"
     gene            14. .1939
                     /gene="mspC"
     CDS             14. .1939
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdG1"
                     /product="GP63"
                     /protein_id="CAD42814.1"
                     /db_xref="GI:21954462"
                     /db_xref="GOA:Q8MNZ2"
                     /db_xref="SPTREMBL:Q8MNZ2"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMHPRVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTADDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWKVTDMVDEICGDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILVVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCGTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVCGSMDYVNCTPGLRVELSXVSSAFEEGGYITCPPYVEVCQAN
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     14. .130
                     /gene="mspC"
     mat_peptide     131. .1822
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 2107;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        764 TACGACCAGCTGGTGACACGTGTCGTCACGCACGAGATGGCGCACGC 810


RESULT 7
LDO495003
LOCUS       LDO495003               2109 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain
            MHOM/KE/1954/LRC-L53.
ACCESSION   AJ495003
VERSION     AJ495003.1  GI:21954457
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2109)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2109
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MHOM/KE/1954/LRC-L53"
                     /db_xref="taxon:5661"
                     /country="Kenya"
     gene            15. .1940
                     /gene="mspC"
     CDS             15. .1940
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdA2"
                     /product="GP63"
                     /protein_id="CAD42812.1"
                     /db_xref="GI:21954458"
                     /db_xref="GOA:Q8MNZ4"
                     /db_xref="SPTREMBL:Q8MNZ4"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVACHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTAEDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWNVTDMVDEICGDFKVPQAHITEGFSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILLVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVRGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQAN
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     15. .131
                     /gene="mspC"
     mat_peptide     132. .1823
                     /gene="mspC"
                     /product="GP63"
ORIGIN
```

```
  Query Match            61.2%;  Score 29.4;  DB 3;  Length 2109;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        765 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 811


RESULT 8
LDO495004
LOCUS       LDO495004               2109 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain
            MHOM/CN/0000/WangJie1.
ACCESSION   AJ495004
VERSION     AJ495004.1  GI:21954459
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2109)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2109
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MHOM/CN/0000/WangJie1"
                     /db_xref="taxon:5661"
                     /country="China"
     gene            15. .1940
                     /gene="mspC"
     CDS             15. .1940
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdF"
                     /product="GP63"
                     /protein_id="CAD42813.1"
                     /db_xref="GI:21954460"
                     /db_xref="GOA:Q8MNZ3"
                     /db_xref="SPTREMBL:Q8MNZ3"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTAEDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWNVTDMVDEICGDFKVPQAHITEGFSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILLVTQMMNIRGKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNESEVTMRCPTSRLMVGTCGIRGYSTPFSLYWQYFTNASLGGYSPFL
                     DYCPFVIGYSDGSCNQDASLAAGFFSAFNVFSDAARCIDGAFRPKNRTAANGYYAGLC
                     ANVRCDTATRTYSVQVRGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGN
                     VQAAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     15. .131
                     /gene="mspC"
     mat_peptide     132. .1823
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match            61.2%;  Score 29.4;  DB 3;  Length 2109;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        765 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 811


RESULT 9
LEIGP63D
LOCUS       LEIGP63D                2892 bp    mRNA    linear   INV 26-APR-1993
DEFINITION  Leishmania donovani chagasi surface protease (log glycoprotein 63,
            GP63) mRNA, complete cds.
ACCESSION   M80672
VERSION     M80672.1  GI:159328
KEYWORDS    log glycoprotein 63; surface protease.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 2892)
  AUTHORS   Ramamoorthy,R., Donelson,J.E., Paetz,K.E., Maybodi,M., Roberts,S.C.
            and Wilson,M.E.
  TITLE     Three distinct RNAs for the surface protease gp63 are
            differentially expressed during development of Leishmania donovani
            chagasi promastigotes to an infectious form
  JOURNAL   J. Biol. Chem. 267 (3), 1888-1895 (1992)
  MEDLINE   92112918
   PUBMED   1370484
COMMENT     Original source text: Leishmania donovani (sub_species chagasi)
            cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .2892
                     /organism="Leishmania donovani"
                     /mol_type="mRNA"
                     /sub_species="chagasi"
                     /db_xref="taxon:5661"
     gene            1. .2892
                     /gene="GP63"
     CDS             114. .1913
                     /gene="GP63"
                     /note="log glycoprotein 63"
                     /codon_start=1
                     /product="surface protease"
                     /protein_id="AAA29238.1"
                     /db_xref="GI:159329"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
                     ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGPFEG
                     ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
                     CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGIIKSYAGLCA
                     NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
                     QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
     sig_peptide     114. .230
                     /gene="GP63"
     mat_peptide     405. .1910
                     /gene="GP63"
                     /product="surface protease"
                     /note="log glycoprotein 63"
     polyA_site      2892
                     /gene="GP63"
ORIGIN

  Query Match            61.2%;  Score 29.4;  DB 3;  Length 2892;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        864 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 910
```

```
RESULT 10
LEIGP63B
LOCUS       LEIGP63B                2966 bp    mRNA    linear   INV 26-APR-1993
DEFINITION  Leishhmania donovani chagasi surface protease (stationary
            glycoprotein 63, GP63) mRNA, complete cds.
ACCESSION   M80669
VERSION     M80669.1  GI:159324
KEYWORDS    stationary glycoprotein 63; surface protease.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 2966)
  AUTHORS   Ramamoorthy,R., Donelson,J.E., Paetz,K.E., Maybodi,M., Roberts,S.C.
            and Wilson,M.E.
  TITLE     Three distinct RNAs for the surface protease gp63 are
            differentially expressed during development of Leishmania donovani
            chagasi promastigotes to an infectious form
  JOURNAL   J. Biol. Chem. 267 (3), 1888-1895 (1992)
  MEDLINE   92112918
   PUBMED   1370484
COMMENT     Original source text: Leishmania donovani (sub_species chagasi)
            cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .2966
                     /organism="Leishmania donovani"
                     /mol_type="mRNA"
                     /sub_species="chagasi"
                     /db_xref="taxon:5661"
     gene            1. .2966
                     /gene="GP63"
     CDS             129. .1928
                     /gene="GP63"
                     /note="stationary glycoprotein 63"
                     /codon_start=1
                     /product="surface protease"
                     /protein_id="AAA29236.1"
                     /db_xref="GI:159325"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
                     ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWAATCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
                     ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
                     CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGQIKSYAGLCA
                     NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
                     QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
     sig_peptide     129. .245
                     /gene="GP63"
     mat_peptide     420. .1925
                     /gene="GP63"
                     /product="surface protease"
                     /note="stationary glycoprotein 63"
     polyA_site      2966
                     /gene="GP63"
ORIGIN

  Query Match             61.2%;  Score 29.4;  DB 3;  Length 2966;
  Best Local Similarity   76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        879 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 925


RESULT 11
LEIGP63A
LOCUS       LEIGP63A                3047 bp    DNA     linear   INV 26-APR-1993
DEFINITION  L.chagasi major surface glycoprotein (gp63) gene, complete cds.
ACCESSION   M28527
VERSION     M28527.1  GI:159322
KEYWORDS    glycoprotein; protease; surface antigen.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 3047)
  AUTHORS   Miller,R.A., Reed,S.G. and Parsons,M.
  TITLE     Leishmania gp63 molecule implicated in cellular adhesion lacks an
            Arg-Gly-Asp sequence
  JOURNAL   Mol. Biochem. Parasitol. 39 (2), 267-274 (1990)
  MEDLINE   90205976
   PUBMED   2320059
COMMENT     Original source text: L.chagasi (isolate MHOM/BR/82/BA-2C1a) DNA,
            clones pLc63-[1 and 2].
            Draft entry and computer-readable sequence for [1] kindly provided
            by M.Parsons, 03-OCT-1989.
FEATURES             Location/Qualifiers
     source          1. .3047
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /sub_species="chagasi"
                     /db_xref="taxon:5661"
     CDS             496. .2295
                     /note="gp63 protein"
                     /codon_start=1
                     /protein_id="AAA29235.1"
                     /db_xref="GI:159323"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
                     ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
                     ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
                     CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGIIKSYAGLCA
                     NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
                     QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
ORIGIN

  Query Match             61.2%;  Score 29.4;  DB 3;  Length 3047;
  Best Local Similarity   76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db       1246 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 1292


RESULT 12
LIU48798
LOCUS       LIU48798                3047 bp    DNA     linear   INV 15-JAN-1998
DEFINITION  Leishmania infantum surface glycoprotein gp63 (gp63) gene, complete
            cds.
ACCESSION   U48798
VERSION     U48798.1  GI:1213329
KEYWORDS    .
SOURCE      Leishmania infantum
  ORGANISM  Leishmania infantum
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 3047)
  AUTHORS   Gonzalez-Aseguinolaza,G., Almazan,F., Rodriguez,J.F., Marquet,A.
            and Larraga,V.
  TITLE     Cloning of the gp63 surface protease of Leishmania infantum.
            Differential post-translational modifications correlated with
            different infective forms
  JOURNAL   Biochim. Biophys. Acta 1361 (1), 92-102 (1997)
  MEDLINE   97388259
   PUBMED   9247093
REFERENCE   2  (bases 1 to 3047)
```

```
  AUTHORS   Gonzalez-Aseguinolaza,G., Almazan,F., Rodriguez-Aguirre,J.F.,
            Marquet,A. and Larraga,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-1996) Vicente Larraga, Proteins Structure, Center
            of Biological Research, Spanish Research Council, Velazquez 144,
            Madrid 28006, Spain
FEATURES             Location/Qualifiers
     source          1. .3047
                     /organism="Leishmania infantum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5671"
     gene            496. .2295
                     /gene="gp63"
     CDS             496. .2295
                     /gene="gp63"
                     /note="protease"
                     /codon_start=1
                     /product="surface glycoprotein gp63"
                     /protein_id="AAB96339.1"
                     /db_xref="GI:1213330"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
                     ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
                     ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAPLSEKCME
                     RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
                     CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGIIKSYAGLCA
                     NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
                     QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 3047;
  Best Local Similarity   76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db       1246 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 1292


RESULT 13
LEIGPAA
LOCUS       LEIGPAA                 3105 bp    DNA     linear   INV 26-APR-1993
DEFINITION  L.donovani.
ACCESSION   M60048
VERSION     M60048.1  GI:159334
KEYWORDS    glycoprotein 63.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 3105)
  AUTHORS   Webb,J.R., Button,L.L. and McMaster,W.R.
  TITLE     Heterogeneity of the genes encoding the major surface glycoprotein
            of Leishmania donovani
  JOURNAL   Mol. Biochem. Parasitol. 48 (2), 173-184 (1991)
  MEDLINE   92107220
   PUBMED   1762629
COMMENT     Original source text: L.donovani DNA.
FEATURES             Location/Qualifiers
     source          1. .3105
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5661"
     gene            101. .1873
                     /gene="gp63"
     CDS             101. .1873
                     /gene="gp63"
                     /codon_start=1
                     /product="glycoprotein 63"
                     /protein_id="AAA29244.1"
                     /db_xref="GI:159335"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLSYVTLGAAPTVVRAANWGALRIAVST
                     EDLTDSAYHCARVGQRISTRDGRFAICTAEDILTDEKRDILVKYLIPQALQLHTERLK
                     VRQVQDKWKVTGMGNEICGHFKVPPAHITDGLSNTDFVMYVASVPSEGDVLAWATTCQ
                     VFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVVFFRDARILESISNV
                     RHKDFDVPVINSSTAVAKAREQYGCGTLEYLEMEDQGGAGSAGSHIKMRNAQDELMAP
                     ASDAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAPLSEKCMEDGITKWPAMF
                     CNENEVTMRCHTGRLSLGVCGLSSSDIPLPPYWQYFTDPLLAGISAFMDYCPVVVPFG
                     DGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTTETVTNSYAGLCANVRCDTATR
                     TYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNVQAAKDGGNA
                     AAGRRGPRAAATALLVAALLAVAL"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 3105;
  Best Local Similarity   76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        821 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 867


RESULT 14
LEIGP63C
LOCUS       LEIGP63C                2621 bp    mRNA    linear   INV 26-APR-1993
DEFINITION  Leishmania donovani chagasi surface protease (constitutive
            glycoprotein 63, GP63) mRNA, complete cds.
ACCESSION   M80671
VERSION     M80671.1  GI:159326
KEYWORDS    constitutive glycoprotein 63; surface protease.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 2621)
  AUTHORS   Ramamoorthy,R., Donelson,J.E., Paetz,K.E., Maybodi,M., Roberts,S.C.
            and Wilson,M.E.
  TITLE     Three distinct RNAs for the surface protease gp63 are
            differentially expressed during development of Leishmania donovani
            chagasi promastigotes to an infectious form
  JOURNAL   J. Biol. Chem. 267 (3), 1888-1895 (1992)
  MEDLINE   92112918
   PUBMED   1370484
COMMENT     Original source text: Leishmania donovani (sub_species chagasi)
            cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .2621
                     /organism="Leishmania donovani"
                     /mol_type="mRNA"
                     /sub_species="chagasi"
                     /db_xref="taxon:5661"
     gene            1. .2621
                     /gene="GP63"
     CDS             132. .2051
                     /gene="GP63"
                     /note="constitutive glycoprotein 63"
                     /codon_start=1
                     /product="surface protease"
                     /protein_id="AAA29237.1"
                     /db_xref="GI:159327"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTAEDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWNVTGMVDEICGDFKVPPAHITEGFSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHVGFSGTFFTEI
                     LLVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKMR
                     NAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCMER
                     NITKWPAMFCNENEVTMRCPTSRLMVGTCGIRGYSTPFSLYWQYFTNASLGLLAFLDY
                     CPFVIGYSDGSCNQDASLAAGFFSAFNVFSDAARCIDGAFRPKNRTAANGYYAGLCAN
                     VRCDTATRTYSVQVRGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQANVK
                     GAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWLLL
                     LTCPWWCCKFGGLPT"
```

```
     sig_peptide     132. .248
                     /gene="GP63"
     mat_peptide     423. .2048
                     /gene="GP63"
                     /product="surface protease"
                     /note="constitutive glycoprotein 63"
     polyA_site      2621
                     /gene="GP63"
ORIGIN

  Query Match             59.2%;  Score 28.4;  DB 3;  Length 2621;
  Best Local Similarity   76.1%;  Pred. No. 6.8;
  Matches   35;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              || || |||||||| || ||||| || || || || ||||| || |
Db        882 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACG 927


RESULT 15
LTU495010
LOCUS       LTU495010               2050 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania turanica mspC gene for GP63, strain
            MRHO/SU/1983/MARZ-051.
ACCESSION   AJ495010
VERSION     AJ495010.1  GI:21954471
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania turanica
  ORGANISM  Leishmania turanica
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2050)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2050
                     /organism="Leishmania turanica"
                     /mol_type="genomic DNA"
                     /strain="MRHO/SU/1983/MARZ-051"
                     /db_xref="taxon:62297"
     gene            15. .1949
                     /gene="mspC"
     CDS             15. .1949
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLtuA1"
                     /product="GP63"
                     /protein_id="CAD42819.1"
                     /db_xref="GI:21954472"
                     /db_xref="GOA:Q8MNY9"
                     /db_xref="SPTREMBL:Q8MNY9"
                     /translation="MSVDSSSTHRRRCVAARLVRLAAAGVAVTVAVGTAAAWAHAGAL
                     QHRCVHDAMQARVRQSVADHHTAPGAVSAVGLPYVTLDAAHTAAAADPRPGSARGVER
                     DVNWGALRIAVSIEDLTDPXYHCVRVGQXVNNHAGAIVTCTAEDILTDEKRDILVKHL
                     IPRALXLHTERLKVQQVQGTWKVTGMVDEICGDFKVPQAHITKGFSNTDFVMYVASVP
                     SEEGVLAWATTCQVFSDGHPAVGVINIPAANIVSRYDQLVTRVVTHEMAHALGFSGPF
                     FEDARIVANISNVRGKNFDVPVINSSTAVAKAREQYGCDTLEYLEVEDQGGAGSAGSH
                     IKMRNAQDELMAPAAAAGYYTALTMAIFQDLGFYQADFSKAEVMLWGRNAGCAFLTDK
                     CMEQNITQWPAMFCNESEDAIRCPTSRLILGTCGIREYETPLPLYWQYFTNASLGGYS
                     PFLDYCPSVIGYAHGXCNQDASSAAELLAAFNVFSDAARCIDGAFTPKNRTAANGYYG
                     GLCANVQCDTATRTYSVQVHGSMDYANCTPGLRVELSTVSNTFEEGGYITCPPYVEVC
                     QGNVKGAKDFDGDSDSSSSAGDDADKAAMERWNDRMAGLATAVTVLLGIVLSLMALVV
                     VWLLLVSCPRWCCKVGGLPT"
     sig_peptide     15. .131
                     /gene="mspC"
     mat_peptide     132. .1832
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match             57.9%;  Score 27.8;  DB 3;  Length 2050;
  Best Local Similarity   74.5%;  Pred. No. 11;
  Matches   35;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || ||||| || || ||||| || || || || ||||| || ||
Db        774 TACGACCAGCTCGTCACGCGTGTCGTCACGCACGAGATGGCGCACGC 820


Search completed: March 18, 2004, 15:02:39
Job time : 246.087 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 30.3584 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	48	100.0	48	3	AAC64621	Aac64621 Leishmani
c	2	48	100.0	78	3	AAC64609	Aac64609 AgfA (SEF
	3	48	100.0	78	3	AAC64610	Aac64610 AgfA (SEF
	4	48	100.0	456	3	AAC64628	Aac64628 AgfA::PT3
	5	48	100.0	456	3	AAC64622	Aac64622 AgfA::PT3
	6	48	100.0	456	3	AAC64626	Aac64626 AgfA::PT3
	7	48	100.0	456	3	AAC64630	Aac64630 AgfA::PT3
	8	48	100.0	456	3	AAC64629	Aac64629 AgfA::PT3
	9	48	100.0	456	3	AAC64625	Aac64625 AgfA::PT3
	10	48	100.0	456	3	AAC64631	Aac64631 AgfA::PT3
	11	48	100.0	456	3	AAC64623	Aac64623 AgfA::PT3
	12	48	100.0	456	3	AAC64627	Aac64627 AgfA::PT3
	13	48	100.0	456	3	AAC64624	Aac64624 AgfA::PT3
	14	43.2	90.0	48	3	AAC64616	Aac64616 S. enteri
	15	40.6	84.6	78	3	AAC64606	Aac64606 SefA (SEF
c	16	40.6	84.6	78	3	AAC64605	Aac64605 SefA (SEF
	17	27.8	57.9	2547	7	ACA48284	Aca48284 Prokaryot
	18	24.6	51.3	1780	2	AAV65244	Aav65244 S. pneumo
	19	24.6	51.3	2544	7	ABX06472	Abx06472 S. pneumo
	20	24.6	51.3	2547	7	ACA49840	Aca49840 Prokaryot
c	21	24.6	51.3	3636	2	AAV52211	Aav52211 Streptoco
	22	24.6	51.3	110000	7	ABS56454_06	Continuation (7 of
	23	23	47.9	420	6	ABQ69493	Abq69493 Listeria
	24	23	47.9	420	6	ABQ67505	Abq67505 Listeria
c	25	23	47.9	436	8	ACH39062	Ach39062 Human foe
c	26	23	47.9	810	5	AAH52245	Aah52245 Human AFP
	27	23	47.9	951	4	ABL21387	Abl21387 Drosophil
c	28	23	47.9	1722	6	ABQ76709	Abq76709 Human mac
	29	23	47.9	2940	2	AAT36490	Aat36490 Lactococc
	30	23	47.9	3333	4	ABL21386	Abl21386 Drosophil
c	31	23	47.9	4850	7	ABQ83879	Abq83879 Human MDD
	32	23	47.9	18949	4	AAK70445	Aak70445 Human imm
c	33	23	47.9	63020	6	ABQ67197_11	Continuation (12 o
c	34	23	47.9	110000	6	ABQ69245_12	Continuation (13 o
c	35	22.8	47.5	1323	3	AAF13368	Aaf13368 Aspergill
	36	22.6	47.1	1767	7	ACA37452	Aca37452 Prokaryot
	37	22.6	47.1	3122	2	AAQ79913	Aaq79913 Lys-amino
	38	22.4	46.7	1014	4	AAF44754	Aaf44754 Nucellin-
	39	22.4	46.7	1014	9	ADB94941	Adb94941 PCD pathw
	40	22.4	46.7	1925	9	ADB94853	Adb94853 PCD pathw
	41	22.2	46.2	912	5	AAH67547	Aah67547 C glutami
	42	22.2	46.2	1035	4	AAF71443	Aaf71443 Corynebac
	43	22.2	46.2	44147	6	ABK84481	Abk84481 Human cDN
	44	22.2	46.2	44147	9	ADD14691	Add14691 Human src
c	45	22.2	46.2	113604	8	ADA42607	Ada42607 Human LQT

ALIGNMENTS

RESULT 1
AAC64621
ID AAC64621 standard; DNA; 48 BP.
XX
AC AAC64621;
XX
DT 26-FEB-2001 (first entry)
XX
DE Leishmania major PT3 epitope DNA sequence SEQ ID NO:9.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Leishmania major.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36345.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 48 BP; 12 A; 11 C; 11 G; 14 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.4e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48

RESULT 2
AAC64609/c
ID AAC64609 standard; DNA; 78 BP.
XX
AC AAC64609;
XX
DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)
XX
DE AgfA (SEF17) recombinant agfA::PT3a generating PCR primer 17-B.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.
XX
OS Salmonella enteritidis.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Example 2; Page 61; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 78 BP; 17 A; 17 C; 21 G; 23 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 9.5e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 48 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 1

RESULT 3
AAC64610
ID AAC64610 standard; DNA; 78 BP.
XX
AC AAC64610;
XX
DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)
XX
DE AgfA (SEF17) recombinant agfA::PT3a generating PCR primer 17-C.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.
XX
OS Salmonella enteritidis.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Example 2; Page 61; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 78 BP; 21 A; 19 C; 17 G; 21 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 9.5e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48

RESULT 4
AAC64628
ID AAC64628 standard; DNA; 456 BP.
XX
AC AAC64628;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36352.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 331 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 378

RESULT 5
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX

PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 382 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 429

RESULT 6
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 196 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243

RESULT 7
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.

```
DR   P-PSDB; AAB36354.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 48;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-09;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 288


RESULT 8
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 48;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-09;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 219


RESULT 9
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
```

XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 127 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 174

RESULT 10
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36355.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 307 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 354

RESULT 11
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.

XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 352 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 399

RESULT 12
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36351.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 262 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 309

RESULT 13
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.

XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 67 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 114

RESULT 14
AAC64616
ID AAC64616 standard; DNA; 48 BP.
XX
AC AAC64616;
XX
DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)
XX
DE S. enteritidis recombinant agfA gene sequence.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Fig 1c; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 48 BP; 14 A; 10 C; 10 G; 14 T; 0 U; 0 Other;

Query Match 90.0%; Score 43.2; DB 3; Length 48;
Best Local Similarity 93.8%; Pred. No. 7.6e-08;
Matches 45; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
|||||||||||| ||||| |||||||||| ||||||||||||||||||
Db 1 TATGATCAGCTCGTTACACGTGTTGTTACACATGAAATGGCACATGCA 48

RESULT 15
AAC64606
ID AAC64606 standard; DNA; 78 BP.
XX
AC AAC64606;
XX

DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)
XX
DE SefA (SEF14) recombinant sefA::PT3 generating PCR primer 14-C.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.
XX
OS Salmonella enteritidis.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Example 2; Page 61; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 78 BP; 16 A; 16 C; 21 G; 25 T; 0 U; 0 Other;

Query Match 84.6%; Score 40.6; DB 3; Length 78;
Best Local Similarity 91.5%; Pred. No. 1e-06;
Matches 43; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              ||||||||||| ||||| |||||||||||| |||||||||||| |||||
Db          1 TATGATCAGCTCGTTACACGTGTTGTTACACATGAAATGGCTCATGC 47
```

Search completed: March 18, 2004, 09:11:54
Job time : 32.3584 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 5.52601 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	24.6	51.3	3636	4	US-08-961-527-78	Sequence 78, Appl
2	22.4	46.7	1014	4	US-09-325-932A-15	Sequence 15, Appl
c 3	22	45.8	5406	4	US-08-961-527-166	Sequence 166, App
c 4	21.8	45.4	7754	4	US-09-634-238-3	Sequence 3, Appli
c 5	21.4	44.6	648	4	US-09-252-991A-11586	Sequence 11586, A
c 6	21.4	44.6	963	4	US-09-252-991A-11645	Sequence 11645, A
7	21.4	44.6	987	4	US-09-252-991A-11794	Sequence 11794, A
c 8	21.4	44.6	1983	4	US-09-453-702B-36	Sequence 36, Appl
9	21.4	44.6	2166	2	US-08-317-401E-3	Sequence 3, Appli
10	21.4	44.6	2877	2	US-08-317-401E-1	Sequence 1, Appli
c 11	21.4	44.6	9541	4	US-08-961-527-132	Sequence 132, App
12	21	43.8	597	4	US-09-543-681A-3167	Sequence 3167, Ap
13	21	43.8	2975	4	US-08-879-337-10	Sequence 10, Appl
14	21	43.8	6587	4	US-08-879-337-3	Sequence 3, Appli
15	21	43.8	100848	4	US-09-596-002-39	Sequence 39, Appl
c 16	20.8	43.3	705	4	US-09-543-681A-1102	Sequence 1102, Ap
17	20.8	43.3	1029	4	US-09-328-352-2205	Sequence 2205, Ap
18	20.6	42.9	1239	4	US-09-328-352-779	Sequence 779, App
19	20.6	42.9	1872	4	US-09-280-116-110	Sequence 110, App
20	20.6	42.9	3366	4	US-09-345-650-2	Sequence 2, Appli
21	20.6	42.9	8334	4	US-09-543-681A-1952	Sequence 1952, Ap
22	20.4	42.5	440	4	US-09-621-976-8341	Sequence 8341, Ap
23	20.4	42.5	753	4	US-09-252-991A-13584	Sequence 13584, A
c 24	20.4	42.5	765	4	US-09-252-991A-13860	Sequence 13860, A
25	20.4	42.5	1404	4	US-09-107-532A-2521	Sequence 2521, Ap
26	20.4	42.5	1608	4	US-09-328-352-2416	Sequence 2416, Ap
27	20.4	42.5	1848	4	US-08-637-670-23	Sequence 23, Appl
c 28	20.2	42.1	1203	4	US-09-107-532A-2612	Sequence 2612, Ap
c 29	20	41.7	562	3	US-09-328-111-468	Sequence 468, App
30	20	41.7	1152	4	US-09-489-039A-2755	Sequence 2755, Ap
31	20	41.7	1280	3	US-08-335-844A-10	Sequence 10, Appl
32	20	41.7	1280	4	US-09-129-366-10	Sequence 10, Appl
33	20	41.7	1293	3	US-08-335-844A-11	Sequence 11, Appl
34	20	41.7	1293	4	US-09-129-366-11	Sequence 11, Appl
35	20	41.7	2472	3	US-08-335-844A-7	Sequence 7, Appli
36	20	41.7	2472	4	US-09-129-366-7	Sequence 7, Appli
37	20	41.7	3213	4	US-09-107-532A-62	Sequence 62, Appl
38	20	41.7	3296	3	US-08-335-844A-14	Sequence 14, Appl
39	20	41.7	3296	4	US-09-129-366-14	Sequence 14, Appl
40	20	41.7	3305	3	US-08-335-844A-8	Sequence 8, Appli
41	20	41.7	3305	4	US-09-129-366-8	Sequence 8, Appli
42	20	41.7	3319	3	US-08-335-844A-15	Sequence 15, Appl
43	20	41.7	3319	4	US-09-129-366-15	Sequence 15, Appl
44	20	41.7	3358	3	US-08-335-844A-20	Sequence 20, Appl
45	20	41.7	3358	4	US-09-129-366-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-08-961-527-78/c
; Sequence 78, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-78

Query Match 51.3%; Score 24.6; DB 4; Length 3636;
Best Local Similarity 76.9%; Pred. No. 1.1;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              ||| |  ||| ||| ||||||  |||||||| |||| ||
Db       2806 TCAACAAGTTGCCCTTGTTGTGGCCCATGAATTGGCTCA 2768
```

```
RESULT 2
US-09-325-932A-15
; Sequence 15, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
;  APPLICANT:  Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION:  Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of forestry plant develc
;  FILE REFERENCE:  1022
;  CURRENT APPLICATION NUMBER: US/09/325,932A
;  CURRENT FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  206
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 15
;  LENGTH: 1014
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-09-325-932A-15

  Query Match             46.7%;  Score 22.4;  DB 4;  Length 1014;
  Best Local Similarity   66.7%;  Pred. No. 6.4;
  Matches   32;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||| |||| || | | | ||  ||| | | || | || || ||| |
Db        439 TATGACCAGCAGGGTTCTCTTGCAGTTTCACCTGCAGTGACAGATGGA 486


RESULT 3
US-08-961-527-166/c
; Sequence 166, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  166:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5406 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-166

  Query Match             45.8%;  Score 22;  DB 4;  Length 5406;
  Best Local Similarity   67.4%;  Pred. No. 15;
  Matches   31;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | |||||||      ||| | |  |||| ||| | |||| | ||||
Db       3442 AAGATCAGCTAACAGCCCTTATGATTACTCATCATATGGAAGATGC 3397


RESULT 4
US-09-634-238-3/c
; Sequence 3, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
;  APPLICANT:  Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka J.
;  APPLICANT:  Bloksberg, Leonard, N.
;  APPLICANT:  Lubbers, Mark W.
;  APPLICANT:  Dekker, James
;  APPLICANT:  Christensson, Anna C.
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul W.
;  APPLICANT:  Reid, Julian R.
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION:  Polynucleotides, materials incorporating
;  TITLE OF INVENTION:  them and methods for using them.
;  FILE REFERENCE:  11000.1043U1
;  CURRENT APPLICATION NUMBER: US/09/634,238
;  CURRENT FILING DATE:  2000-08-08
;  NUMBER OF SEQ ID NOS:  422
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 7754
;   TYPE: DNA
;   ORGANISM: Lactobacillus rhamnosus
US-09-634-238-3

  Query Match             45.4%;  Score 21.8;  DB 4;  Length 7754;
  Best Local Similarity   70.7%;  Pred. No. 20;
  Matches   29;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          5 ATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | |||| |||| |    ||| || | |||||| |||| |||
Db       4023 AGCAGCGGGTTGCTACCGTTATTGCGCATGAACTGGCGCAT 3983


RESULT 5
US-09-252-991A-11586/c
; Sequence 11586, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11586
;   LENGTH: 648
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11586

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 648;
  Best Local Similarity   71.8%;  Pred. No. 14;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              ||||  || | | |  |||||||| |||||| | ||| |
```

```
Db        296 GATCGCCTCGATCCTGGTGTTGTTTCCCATGCACTGGAA 258


RESULT 6
US-09-252-991A-11645/c
; Sequence 11645, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11645
;   LENGTH: 963
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11645

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 963;
  Best Local Similarity   71.8%;  Pred. No. 16;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              ||||  || | | |  |||||||| |||||| | ||| |
Db        316 GATCGCCTCGATCCTGGTGTTGTTTCCCATGCACTGGAA 278


RESULT 7
US-09-252-991A-11794
; Sequence 11794, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11794
;   LENGTH: 987
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11794

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 987;
  Best Local Similarity   71.8%;  Pred. No. 16;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              ||||  || | | |  |||||||| |||||| | ||| |
Db        681 GATCGCCTCGATCCTGGTGTTGTTTCCCATGCACTGGAA 719


RESULT 8
US-09-453-702B-36/c
; Sequence 36, Application US/09453702B
; Patent No. 6365723
;  GENERAL INFORMATION:
;       APPLICANT: Blattner, Frederick R.
;                  Burland,     Valerie
;                  Perna,  Nicole T.
;                  Plunkett, Guy
;                  Welch,  Rod
;       TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;       NUMBER OF SEQUENCES: 265
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Quarles & Brady
;            STREET: 1 South Pinckney Street
;            CITY: Madison
;            STATE: WI
;            COUNTRY: US
;            ZIP: 53701-2113
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: Word Perfect 8.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/453,702B
;            FILING DATE: 03-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/110,955
;            FILING DATE: 04-DEC-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Seay, Nicholas J.
;            REGISTRATION NUMBER: 27386
;            REFERENCE/DOCKET NUMBER: 960296.95017
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (608) 251-5000
;            TELEFAX: (608) 251-9166
;  INFORMATION FOR SEQ ID NO: 36:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1983
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA  (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-453-702B-36

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 1983;
  Best Local Similarity   66.0%;  Pred. No. 20;
  Matches   31;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||||||  |||     | | | ||||||| ||||||   | ||| |
Db        372 ATGATCAATTGGACCAACTTTTGGTTACCCGTGAAATTCAAGATGAA 326


RESULT 9
US-08-317-401E-3
; Sequence 3, Application US/08317401E
; Patent No. 5922561
;  GENERAL INFORMATION:
;    APPLICANT:  Thompson, Sheryl Ann
;    APPLICANT:  Yaver, Debbie Sue
;    TITLE OF INVENTION:  GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
;    TITLE OF INVENTION:  ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,401E
```

```
;      FILING DATE:  03-October-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4248.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2166 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
US-08-317-401E-3

  Query Match             44.6%;  Score 21.4;  DB 2;  Length 2166;
  Best Local Similarity   80.6%;  Pred. No. 20;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATG 34
              ||| |||||||| |||  |||| | ||||||
Db        326 GATGAGCTGGTTTCCCTGGTTGATCCCCATG 356


RESULT 10
US-08-317-401E-1
; Sequence 1, Application US/08317401E
; Patent No. 5922561
;  GENERAL INFORMATION:
;    APPLICANT:  Thompson, Sheryl Ann
;    APPLICANT:  Yaver, Debbie Sue
;    TITLE OF INVENTION:  GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
;    TITLE OF INVENTION:  ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,401E
;      FILING DATE:  03-October-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4248.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2877 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join (126..203, 253..1776)
US-08-317-401E-1

  Query Match             44.6%;  Score 21.4;  DB 2;  Length 2877;
  Best Local Similarity   80.6%;  Pred. No. 22;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATG 34
              ||| |||||||| |||  |||| | ||||||
Db        418 GATGAGCTGGTTTCCCTGGTTGATCCCCATG 448


RESULT 11
US-08-961-527-132/c
; Sequence 132, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  132:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9541 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-132

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 9541;
  Best Local Similarity   71.8%;  Pred. No. 31;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          7 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              ||| || ||| |||||  | || | || |||||| | ||
Db       6099 CAGTTGATTAGCCGTGGGGCTATCAATAAAATGGGAAAT 6061


RESULT 12
US-09-543-681A-3167
; Sequence 3167, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABII
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
```

```
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 3167
;   LENGTH: 597
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-3167

  Query Match             43.8%;  Score 21;  DB 4;  Length 597;
  Best Local Similarity   73.0%;  Pred. No. 21;
  Matches   27;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         12 GGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||     ||||| ||| | ||||||| |||| |||
Db        375 GGTTGGAACTGTTGCTACACTTGAAATGACACAAGCA 411


RESULT 13
US-08-879-337-10
; Sequence 10, Application US/08879337B
; Patent No. 6639130
; GENERAL INFORMATION:
;  APPLICANT: Jang, Jyan-Chyun
;  APPLICANT: Sheen, Jen
;  TITLE OF INVENTION: PLANT STEROL REDUCTASES AND USES THEREOF
;  FILE REFERENCE: 00786/338001
;  CURRENT APPLICATION NUMBER: US/08/879,337B
;  CURRENT FILING DATE: 1997-06-20
;  EARLIER APPLICATION NUMBER: 60/022,086
;  EARLIER FILING DATE: 1996-06-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 2975
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)...(2975)
;   OTHER INFORMATION: N can be any nucleotide.
US-08-879-337-10

  Query Match             43.8%;  Score 21;  DB 4;  Length 2975;
  Best Local Similarity   66.7%;  Pred. No. 32;
  Matches   30;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | | ||||| ||| |  |    | |||||| ||| ||||| | ||
Db        310 TTTTATCAGATGGCTCTCAACTTCGTTACCGATGCAATGGTATAT 354


RESULT 14
US-08-879-337-3
; Sequence 3, Application US/08879337B
; Patent No. 6639130
; GENERAL INFORMATION:
;  APPLICANT: Jang, Jyan-Chyun
;  APPLICANT: Sheen, Jen
;  TITLE OF INVENTION: PLANT STEROL REDUCTASES AND USES THEREOF
;  FILE REFERENCE: 00786/338001
;  CURRENT APPLICATION NUMBER: US/08/879,337B
;  CURRENT FILING DATE: 1997-06-20
;  EARLIER APPLICATION NUMBER: 60/022,086
;  EARLIER FILING DATE: 1996-06-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 6587
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)...(6587)
;   OTHER INFORMATION: N can be any nucleotide.
US-08-879-337-3

  Query Match             43.8%;  Score 21;  DB 4;  Length 6587;
  Best Local Similarity   66.7%;  Pred. No. 40;
  Matches   30;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | | ||||| ||| |  |    | |||||| ||| ||||| | ||
Db       3664 TTTTATCAGATGGCTCTCAACTTCGTTACCGATGCAATGGTATAT 3708


RESULT 15
US-09-596-002-39
; Sequence 39, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:      Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:      PERL Program
; SEQ ID NO 39
;   LENGTH: 100848
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte template ID No. 6632636 39
;  PUBLICATION INFORMATION:
US-09-596-002-39

  Query Match             43.8%;  Score 21;  DB 4;  Length 100848;
  Best Local Similarity   66.7%;  Pred. No. 86;
  Matches   30;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||  |||  | || | |    ||| |||| |||||||||| | |
Db      73642 GATGTGCTATTGACACATACATTTATCCATCAAATGGCACAAGAA 73686


Search completed: March 18, 2004, 20:20:10
Job time : 6.52601 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 26.3931 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	27.8	57.9	2547	12	US-10-282-122A-36154	Sequence 36154, A
	2	24.6	51.3	2547	12	US-10-282-122A-37710	Sequence 37710, A
c	3	24.6	51.3	3636	12	US-10-158-844-78	Sequence 78, Appl
c	4	24	50.0	409	12	US-10-424-599-119735	Sequence 119735,
	5	23.2	48.3	720	12	US-10-425-114-6010	Sequence 6010, Ap
	6	23	47.9	420	15	US-10-398-221-318	Sequence 318, App
	7	23	47.9	420	15	US-10-398-221-2306	Sequence 2306, Ap
c	8	23	47.9	436	10	US-09-918-995-26274	Sequence 26274, A
c	9	23	47.9	1163020	15	US-10-398-221-10	Sequence 10, Appl
c	10	23	47.9	3011208	15	US-10-398-221-2058	Sequence 2058, Ap
	11	22.6	47.1	1767	12	US-10-282-122A-25322	Sequence 25322, A
	12	22.6	47.1	3984	15	US-10-369-493-42554	Sequence 42554, A
	13	22.4	46.7	1014	14	US-10-219-220-15	Sequence 15, Appl
	14	22.4	46.7	1925	14	US-10-219-220-217	Sequence 217, App
	15	22.2	46.2	912	9	US-09-738-626-2582	Sequence 2582, Ap
c	16	22.2	46.2	2370	15	US-10-369-493-32121	Sequence 32121, A
c	17	22.2	46.2	2487	15	US-10-369-493-44598	Sequence 44598, A
c	18	22.2	46.2	2574	15	US-10-369-493-34295	Sequence 34295, A
c	19	22.2	46.2	113604	14	US-10-227-195A-1	Sequence 1, Appli
c	20	22.2	46.2	113604	14	US-10-227-195A-2	Sequence 2, Appli
c	21	22.2	46.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
c	22	22	45.8	145	9	US-09-864-761-24629	Sequence 24629, A
c	23	22	45.8	480	14	US-10-226-956-320	Sequence 320, App
c	24	22	45.8	557	9	US-09-864-761-7925	Sequence 7925, Ap
c	25	22	45.8	1791	9	US-09-822-830A-241	Sequence 241, App
c	26	22	45.8	5406	12	US-10-158-844-166	Sequence 166, App
	27	21.8	45.4	1470	12	US-10-424-599-56999	Sequence 56999, A
	28	21.8	45.4	1510	12	US-10-425-114-36154	Sequence 36154, A
	29	21.8	45.4	1646	12	US-10-425-114-31422	Sequence 31422, A
	30	21.8	45.4	2394	12	US-10-425-114-32979	Sequence 32979, A
	31	21.8	45.4	3328	12	US-10-425-114-34623	Sequence 34623, A
c	32	21.8	45.4	7754	15	US-10-264-213-3	Sequence 3, Appli
c	33	21.8	45.4	8495	13	US-10-108-605-356	Sequence 356, App
c	34	21.6	45.0	633	9	US-09-917-800A-1050	Sequence 1050, Ap
	35	21.6	45.0	729	9	US-09-738-626-1352	Sequence 1352, Ap
	36	21.6	45.0	852	12	US-10-627-476-367	Sequence 367, App
c	37	21.6	45.0	1580	12	US-10-425-114-21821	Sequence 21821, A
	38	21.6	45.0	1746	15	US-10-260-238-578	Sequence 578, App
	39	21.6	45.0	2853	12	US-10-425-114-13682	Sequence 13682, A
c	40	21.6	45.0	1691139	14	US-10-067-514-1	Sequence 1, Appli
c	41	21.6	45.0	1691139	15	US-10-419-723-1	Sequence 1, Appli
	42	21.6	45.0	3309400	9	US-09-738-626-1	Sequence 1, Appli
	43	21.4	44.6	1011	12	US-10-282-122A-30293	Sequence 30293, A
	44	21.4	44.6	1179	10	US-09-769-744A-119	Sequence 119, App
c	45	21.4	44.6	1983	14	US-10-114-170-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-10-282-122A-36154
; Sequence 36154, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

```
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36154
;   LENGTH: 2547
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
US-10-282-122A-36154

  Query Match             57.9%;  Score 27.8;  DB 12;  Length 2547;
  Best Local Similarity   82.1%;  Pred. No. 0.27;
  Matches   32;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          7 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              |||| |||| | | ||||||| |||||||| |||| |||
Db        850 CAGCAGGTTGCACTTGTTGTTGCCCATGAACTGGCGCAT 888


RESULT 2
US-10-282-122A-37710
; Sequence 37710, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37710
;   LENGTH: 2547
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-10-282-122A-37710

  Query Match             51.3%;  Score 24.6;  DB 12;  Length 2547;
  Best Local Similarity   76.9%;  Pred. No. 6;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              ||| |  ||| ||| |||||| ||||||||| |||| ||
Db        849 TCAACAAGTTGCCCTTGTTGTGGCCCATGAATTGGCTCA 887


RESULT 3
US-10-158-844-78/c
; Sequence 78, Application US/10158844
; Publication No. US20040029118A1
;  GENERAL INFORMATION:
;       APPLICANT: Kunsch et al.
;       TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
;       NUMBER OF SEQUENCES: 391
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD-R
;            COMPUTER: Dell Latitude Pentium 3
;            OPERATING SYSTEM: Windows 98
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/158,844
;            FILING DATE: 03-Jun-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 08/961,527
;            FILING DATE: 1997-10-30
;            APPLICATION NUMBER: US 60/029,960
;            FILING DATE: 1996-10-31
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Hyman, Mark J.
;            REGISTRATION NUMBER: 46,789
;            REFERENCE/DOCKET NUMBER: PB340P1D1
;   INFORMATION FOR SEQ ID NO: 78:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 3636 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-10-158-844-78

  Query Match             51.3%;  Score 24.6;  DB 12;  Length 3636;
  Best Local Similarity   76.9%;  Pred. No. 6.6;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              ||| |  ||| ||| |||||| ||||||||| |||| ||
Db       2806 TCAACAAGTTGCCCTTGTTGTGGCCCATGAATTGGCTCA 2768


RESULT 4
US-10-424-599-119735/c
; Sequence 119735, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
```

```
; SEQ ID NO 119735
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_79129C.1
US-10-424-599-119735

  Query Match             50.0%;  Score 24;  DB 12;  Length 409;
  Best Local Similarity   68.8%;  Pred. No. 6.2;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              || | | | | | || |||||||| |  |||  |||| ||| ||| ||
Db        340 TACGGTGACCAGCTTGCCCGTGTTCTCGCCCGAGAAAAGGCGCATCCA 293


RESULT 5
US-10-425-114-6010
; Sequence 6010, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 6010
;   LENGTH: 720
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700550972_FLI
US-10-425-114-6010

  Query Match             48.3%;  Score 23.2;  DB 12;  Length 720;
  Best Local Similarity   70.5%;  Pred. No. 16;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              ||| |||||||||| ||  | || || ||  ||| |  || |||
Db        285 ATGTTCAGCTGGTTCCCGATCTTTTTCCCATTGAGAAAGCCCAT 328


RESULT 6
US-10-398-221-318
; Sequence 318, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 318
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-318

  Query Match             47.9%;  Score 23;  DB 15;  Length 420;
  Best Local Similarity   68.1%;  Pred. No. 16;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | |||| | | |||||    ||| ||| | |||||| | |  |||||
Db        175 TCTGATTATCAGGTTAAGTATGTAGTTGCTCATGAACTAGGGCATGC 221


RESULT 7
US-10-398-221-2306
; Sequence 2306, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2306
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2306

  Query Match             47.9%;  Score 23;  DB 15;  Length 420;
  Best Local Similarity   68.1%;  Pred. No. 16;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | |||| | | |||||    ||| ||| | |||||| | |  |||||
Db        175 TCTGATTATCAGGTTAAGTATGTAGTTGCTCATGAACTAGGGCATGC 221


RESULT 8
US-09-918-995-26274/c
; Sequence 26274, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 26274
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(436)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26274

  Query Match             47.9%;  Score 23;  DB 10;  Length 436;
  Best Local Similarity   68.1%;  Pred. No. 17;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              | ||||| ||||||| | | || |||  |  |||| | || ||  ||
```

```
Db        347 ACGATCACCTGGTTATCTGGGTGGTTGACAGTGAAGTAGCCCACACA 301


RESULT 9
US-10-398-221-10/c
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;   LENGTH: 1163020
;   TYPE: DNA
;   ORGANISM: Listeria innocua
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(end)
;   OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

  Query Match             47.9%;  Score 23;  DB 15;  Length 1163020;
  Best Local Similarity   68.1%;  Pred. No. 1.7e+02;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | ||||  | | |||||    ||| ||| | |||||| | |  |||||
Db    1158751 TCTGATTATCAGGTTAAGTATGTAGTTGCTCATGAACTAGGGCATGC 1158705


RESULT 10
US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
;   LENGTH: 3011208
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2058

  Query Match             47.9%;  Score 23;  DB 15;  Length 3011208;
  Best Local Similarity   68.1%;  Pred. No. 2.1e+02;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | ||||  | | |||||    ||| ||| | |||||| | |  |||||
Db    1249014 TCTGATTATCAGGTTAAGTATGTAGTTGCTCATGAACTAGGGCATGC 1248968


RESULT 11
US-10-282-122A-25322
; Sequence 25322, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25322
;   LENGTH: 1767
;   TYPE: DNA
;   ORGANISM: Legionella pneumophila
US-10-282-122A-25322

  Query Match             47.1%;  Score 22.6;  DB 12;  Length 1767;
  Best Local Similarity   68.9%;  Pred. No. 37;
  Matches   31;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              |||   |||  ||||| | |||  |||   |||| ||| ||||||
Db        929 ATGTCGAGCGAGTTACTCATGTGATTAATTATGATATGCCACATG 973


RESULT 12
US-10-369-493-42554
; Sequence 42554, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
```

```
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42554
;   LENGTH: 3984
;   TYPE: DNA
;   ORGANISM: Anabaena PCC7120
US-10-369-493-42554

  Query Match             47.1%;  Score 22.6;  DB 15;  Length 3984;
  Best Local Similarity   75.7%;  Pred. No. 47;
  Matches   28;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              ||| || |||||| ||| || | ||||| |  |||||
Db        843 TCACCTAGTTACCGGTGATGATGCCCATTATGTGGCA 879


RESULT 13
US-10-219-220-15
; Sequence 15, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
;  APPLICANT: Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION: Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of plant development
;  FILE REFERENCE: 11000.1022c1
;  CURRENT APPLICATION NUMBER: US/10/219,220
;  CURRENT FILING DATE:  2002-08-14
;  PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
;  PRIOR FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 290
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15
;   LENGTH: 1014
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-10-219-220-15

  Query Match             46.7%;  Score 22.4;  DB 14;  Length 1014;
  Best Local Similarity   66.7%;  Pred. No. 38;
  Matches   32;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||| |||| || | | | ||  ||| | | || | || || ||| |
Db        439 TATGACCAGCAGGGTTCTCTTGCAGTTTCACCTGCAGTGACAGATGGA 486


RESULT 14
US-10-219-220-217
; Sequence 217, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
;  APPLICANT: Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION: Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of plant development
;  FILE REFERENCE: 11000.1022c1
;  CURRENT APPLICATION NUMBER: US/10/219,220
;  CURRENT FILING DATE:  2002-08-14
;  PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
;  PRIOR FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 290
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 217
;   LENGTH: 1925
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-10-219-220-217

  Query Match             46.7%;  Score 22.4;  DB 14;  Length 1925;
  Best Local Similarity   66.7%;  Pred. No. 46;
  Matches   32;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||| |||| || | | | ||  ||| | | || | || || ||| |
Db        980 TATGACCAGCAGGGTTCTCTTGCAGTTTCACCTGCAGTGACAGATGGA 1027


RESULT 15
US-09-738-626-2582
; Sequence 2582, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2582
;   LENGTH: 912
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-2582

  Query Match             46.2%;  Score 22.2;  DB 9;  Length 912;
  Best Local Similarity   69.8%;  Pred. No. 44;
  Matches   30;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          3 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              ||||  ||||||  |  |||| |||   | |||| | ||||| |
Db        180 TGATGCGCTGGTCAAGCGTGGTGTCGTCAATGAGAAGGCACGT 222


Search completed: March 19, 2004, 03:49:08
Job time : 39.3931 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 279.029 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	100.0	48	23	US-09-543-407-9	Sequence 9, Appli
c 2	48	100.0	78	23	US-09-543-407-52	Sequence 52, Appl
3	48	100.0	78	23	US-09-543-407-53	Sequence 53, Appl
4	48	100.0	456	23	US-09-543-407-11	Sequence 11, Appl
5	48	100.0	456	23	US-09-543-407-13	Sequence 13, Appl
6	48	100.0	456	23	US-09-543-407-15	Sequence 15, Appl
7	48	100.0	456	23	US-09-543-407-17	Sequence 17, Appl
8	48	100.0	456	23	US-09-543-407-19	Sequence 19, Appl
9	48	100.0	456	23	US-09-543-407-21	Sequence 21, Appl
10	48	100.0	456	23	US-09-543-407-23	Sequence 23, Appl
11	48	100.0	456	23	US-09-543-407-25	Sequence 25, Appl
12	48	100.0	456	23	US-09-543-407-27	Sequence 27, Appl
13	48	100.0	456	23	US-09-543-407-29	Sequence 29, Appl
c 14	40.6	84.6	78	23	US-09-543-407-48	Sequence 48, Appl
15	40.6	84.6	78	23	US-09-543-407-49	Sequence 49, Appl
16	29.4	61.2	3105	3	US-07-965-673A-9	Sequence 9, Appli
17	27.8	57.9	2547	48	US-10-282-122A-36154	Sequence 36154, A
c 18	26.8	55.8	107223	21	US-09-528-237A-850	Sequence 850, App
19	24.6	51.3	1780	1	PCT-US97-22578-44	Sequence 44, Appl
20	24.6	51.3	2448	15	US-09-107-433-2456	Sequence 2456, Ap
21	24.6	51.3	2448	53	US-10-617-320-2456	Sequence 2456, Ap
22	24.6	51.3	2544	1	PCT-US03-27401-93	Sequence 93, Appl
23	24.6	51.3	2544	52	US-10-472-928-1519	Sequence 1519, Ap
24	24.6	51.3	2547	24	US-09-583-110-77	Sequence 77, Appl
25	24.6	51.3	2547	48	US-10-282-122A-37710	Sequence 37710, A
26	24.6	51.3	2547	53	US-10-640-833-77	Sequence 77, Appl
27	24.6	51.3	2867	62	US-60-068-175-437	Sequence 437, App
c 28	24.6	51.3	3636	46	US-10-158-844-78	Sequence 78, Appl
29	24.6	51.3	3657	58	US-60-029-960-47	Sequence 47, Appl
30	24.6	51.3	23131	62	US-60-061-998-613	Sequence 613, App
31	24.6	51.3	2162598	52	US-10-472-928-4979	Sequence 4979, Ap
c 32	24	50.0	409	24	US-09-565-240-34030	Sequence 34030, A
c 33	24	50.0	409	27	US-09-654-617-77662	Sequence 77662, A
c 34	24	50.0	409	29	US-09-684-016-77662	Sequence 77662, A
c 35	24	50.0	409	51	US-10-424-599-119735	Sequence 119735,
36	24	50.0	2023	26	US-09-620-392-69478	Sequence 69478, A
c 37	24	50.0	5823	51	US-10-437-963-69258	Sequence 69258, A
c 38	24	50.0	12979	30	US-09-702-134-5750	Sequence 5750, Ap
c 39	24	50.0	12979	33	US-09-815-264-58710	Sequence 58710, A
c 40	24	50.0	13038	26	US-09-620-392-30380	Sequence 30380, A
41	23.6	49.2	390	21	US-09-521-640-44946	Sequence 44946, A
c 42	23.4	48.8	119	35	US-09-865-439A-100294	Sequence 100294,
c 43	23.4	48.8	119	76	US-60-207-458-144215	Sequence 144215,
44	23.2	48.3	310	27	US-09-654-617-237766	Sequence 237766,
45	23.2	48.3	310	29	US-09-684-016-237766	Sequence 237766,

ALIGNMENTS

RESULT 1
US-09-543-407-9
; Sequence 9, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 48
; TYPE: DNA
; ORGANISM: Leishmania major

US-09-543-407-9

Query Match 100.0%; Score 48; DB 23; Length 48;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48

RESULT 2
US-09-543-407-52/c
; Sequence 52, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PCR primer
US-09-543-407-52

Query Match 100.0%; Score 48; DB 23; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 48 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 1

RESULT 3
US-09-543-407-53
; Sequence 53, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PCR primer
US-09-543-407-53

Query Match 100.0%; Score 48; DB 23; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48

```
RESULT 4
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match            100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 429


RESULT 5
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match            100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 399


RESULT 6
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match            100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 114


RESULT 7
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match            100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 174


RESULT 8
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
```

```
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243


RESULT 9
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 309


RESULT 10
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 378


RESULT 11
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 219


RESULT 12
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
```

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match            100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 288


RESULT 13
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

  Query Match            100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 354


RESULT 14
US-09-543-407-48/c
; Sequence 48, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PCR primer
US-09-543-407-48

  Query Match            84.6%;  Score 40.6;  DB 23;  Length 78;
  Best Local Similarity  91.5%;  Pred. No. 1.1e-05;
  Matches   43;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              |||||||||| ||||| |||||||||| |||||||||| |||||
Db         48 TATGATCAGCTCGTTACACGTGTTGTTACACATGAAATGGCTCATGC 2


RESULT 15
US-09-543-407-49
; Sequence 49, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PCR primer
US-09-543-407-49

  Query Match            84.6%;  Score 40.6;  DB 23;  Length 78;
  Best Local Similarity  91.5%;  Pred. No. 1.1e-05;
  Matches   43;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              |||||||||| ||||| |||||||||| |||||||||| |||||
Db          1 TATGATCAGCTCGTTACACGTGTTGTTACACATGAAATGGCTCATGC 47


Search completed: March 19, 2004, 03:01:48
Job time : 282.029 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 5.69942 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23.2	48.3	435	6	US-10-767-795-1394	Sequence 1394, Ap
	2	23.2	48.3	539	6	US-10-767-795-1392	Sequence 1392, Ap
c	3	23.2	48.3	545	6	US-10-767-701-29396	Sequence 29396, A
	4	23.2	48.3	743	6	US-10-767-701-11493	Sequence 11493, A
	5	22.2	46.2	1035	6	US-10-781-014-167	Sequence 167, App
	6	22	45.8	808	6	US-10-779-543-5864	Sequence 5864, Ap
	7	21.6	45.0	201	6	US-10-767-471-34311	Sequence 34311, A
	8	21.6	45.0	2895	6	US-10-767-701-14458	Sequence 14458, A
	9	21.6	45.0	546025	6	US-10-767-471-10717	Sequence 10717, A
c	10	21.4	44.6	607	6	US-10-767-701-31393	Sequence 31393, A
c	11	21	43.8	26371	6	US-10-765-790-41	Sequence 41, Appl
	12	20.8	43.3	175737	6	US-10-783-271-34	Sequence 34, Appl
c	13	20.6	42.9	421	6	US-10-779-543-15761	Sequence 15761, A
	14	20.6	42.9	905	6	US-10-781-469-49	Sequence 49, Appl
	15	20.6	42.9	198161	6	US-10-775-169-52	Sequence 52, Appl
c	16	20.4	42.5	201	6	US-10-767-471-11899	Sequence 11899, A
	17	20.4	42.5	1404	6	US-10-417-884A-2521	Sequence 2521, Ap
c	18	20.4	42.5	318752	6	US-10-767-471-10553	Sequence 10553, A
c	19	20.2	42.1	532	6	US-10-767-701-30765	Sequence 30765, A
	20	20.2	42.1	779	6	US-10-781-014-169	Sequence 169, App
c	21	20.2	42.1	1203	6	US-10-417-884A-2612	Sequence 2612, Ap
	22	20.2	42.1	4503	6	US-10-417-375A-31	Sequence 31, Appl
	23	20.2	42.1	106776	6	US-10-417-375A-30	Sequence 30, Appl
c	24	20	41.7	596	6	US-10-767-701-657	Sequence 657, App
	25	20	41.7	746	6	US-10-779-543-5903	Sequence 5903, Ap
c	26	20	41.7	863	6	US-10-767-701-8642	Sequence 8642, Ap
	27	20	41.7	3213	6	US-10-417-884A-62	Sequence 62, Appl
	28	20	41.7	36135	6	US-10-100-683-11827	Sequence 11827, A
	29	19.8	41.2	390	6	US-10-779-543-12187	Sequence 12187, A
c	30	19.8	41.2	594	6	US-10-417-884A-380	Sequence 380, App
	31	19.8	41.2	600	7	US-60-545-213-391	Sequence 391, App
	32	19.8	41.2	600	7	US-60-545-213-392	Sequence 392, App
	33	19.8	41.2	600	7	US-60-545-213-542	Sequence 542, App
	34	19.8	41.2	600	7	US-60-545-213-4663	Sequence 4663, Ap
	35	19.8	41.2	600	7	US-60-545-213-4664	Sequence 4664, Ap
	36	19.8	41.2	600	7	US-60-545-213-4814	Sequence 4814, Ap
c	37	19.8	41.2	757	6	US-10-100-683-5410	Sequence 5410, Ap
	38	19.8	41.2	844	6	US-10-767-701-10879	Sequence 10879, A
	39	19.8	41.2	7652	1	PCT-US03-31941A-1	Sequence 1, Appli
c	40	19.8	41.2	8684	1	PCT-US03-31941A-2	Sequence 2, Appli
c	41	19.8	41.2	10011	1	PCT-US03-31941A-19	Sequence 19, Appl
	42	19.6	40.8	201	6	US-10-767-471-34488	Sequence 34488, A
	43	19.6	40.8	201	6	US-10-767-471-34490	Sequence 34490, A
	44	19.6	40.8	201	6	US-10-767-471-34556	Sequence 34556, A
	45	19.6	40.8	1013	6	US-10-767-701-12512	Sequence 12512, A

ALIGNMENTS

RESULT 1
US-10-767-795-1394
; Sequence 1394, Application US/10767795
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 1394
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: uC-gsronu33B172b05b1
US-10-767-795-1394

Query Match 48.3%; Score 23.2; DB 6; Length 435;
Best Local Similarity 70.5%; Pred. No. 1.7;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy          3 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              |||| | ||  | || ||| | |||||| | |||   |||||||
Db         73 TGATAAACTCATAACTCGTTTGGTTACCAAGGAAGAAGCACATG 116
```

RESULT 2
US-10-767-795-1392
; Sequence 1392, Application US/10767795
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 1392
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: uC-gsronu33B027f06b1

```
US-10-767-795-1392

  Query Match             48.3%;  Score 23.2;  DB 6;  Length 539;
  Best Local Similarity   70.5%;  Pred. No. 1.7;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          3 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              |||| | ||  | || ||| | |||||| | |||   |||||||
Db         29 TGATAAACTCATAACTCGTTTGGTTACCAAGGAAGAAGCACATG 72


RESULT 3
US-10-767-701-29396/c
; Sequence 29396, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 29396
;   LENGTH: 545
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 9299890
US-10-767-701-29396

  Query Match             48.3%;  Score 23.2;  DB 6;  Length 545;
  Best Local Similarity   77.8%;  Pred. No. 1.7;
  Matches   28;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          5 ATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGG 40
              |||||||||| |||| ||  | |||||||||   ||
Db        532 ATCAGCTGGTGACCCATGCAGATACCCATGACGGGG 497


RESULT 4
US-10-767-701-11493
; Sequence 11493, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 11493
;   LENGTH: 743
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS45512_1
US-10-767-701-11493

  Query Match             48.3%;  Score 23.2;  DB 6;  Length 743;
  Best Local Similarity   70.5%;  Pred. No. 1.9;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              ||| | |||||||| ||| | || || ||  ||| |  || |||
Db        269 ATGTTTAGCTGGTTCCCCATCTTCTTCCCATTGAGAAAGCCCAT 312


RESULT 5
US-10-781-014-167
; Sequence 167, Application US/10781014
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
;  TITLE OF INVENTION:  INVOLVED IN CARBON METABOLISM AND ENERGY
;  TITLE OF INVENTION:  PRODUCTION
;  FILE REFERENCE: BGI-126CPCN
;  CURRENT APPLICATION NUMBER: US/10/781,014
;  CURRENT FILING DATE:  2004-02-17
;  PRIOR APPLICATION NUMBER: US 09/602,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/143,208
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: 60/151,572
;  PRIOR FILING DATE: 1999-08-31
;  PRIOR APPLICATION NUMBER: DE 19931412.8
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931413.6
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931424.1
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931428.4
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 167
;   LENGTH: 1035
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101)..(1012)
;   OTHER INFORMATION: RXN02920
US-10-781-014-167

  Query Match             46.2%;  Score 22.2;  DB 6;  Length 1035;
  Best Local Similarity   69.8%;  Pred. No. 5.2;
  Matches   30;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          3 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              ||||  |||||| |  |||| |||   | |||| | ||||| |
Db        280 TGATGCGCTGGTCAAGCGTGGTGTCGTCAATGAGAAGGCACGT 322


RESULT 6
US-10-779-543-5864
; Sequence 5864, Application US/10779543
; GENERAL INFORMATION:
;  APPLICANT: Williams et al
;  TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
;  TITLE OF INVENTION:  IN CANCEROUS CELLS AND THEIR METHODS OF USE II
;  FILE REFERENCE: 2300-21302
;  CURRENT APPLICATION NUMBER: US/10/779,543
;  CURRENT FILING DATE:  2004-02-12
;  PRIOR APPLICATION NUMBER: 10/076,555
;  PRIOR FILING DATE: 2002-02-15
;  PRIOR APPLICATION NUMBER: 09/217,471
;  PRIOR FILING DATE: 1998-12-21
;  PRIOR APPLICATION NUMBER: 60/068,755
;  PRIOR FILING DATE: 1997-12-23
;  PRIOR APPLICATION NUMBER: 60/080,664
;  PRIOR FILING DATE: 1998-04-03
;  PRIOR APPLICATION NUMBER: 60/105,234
```

```
;  PRIOR FILING DATE: 1998-10-21
;  PRIOR APPLICATION NUMBER: 09/297,648
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: PCT/US99/01619
;  PRIOR FILING DATE: 1999-01-28
;  PRIOR APPLICATION NUMBER: 60/072,910
;  PRIOR FILING DATE: 1998-01-28
;  PRIOR APPLICATION NUMBER: 60/075,954
;  PRIOR FILING DATE: 1998-02-24
;  PRIOR APPLICATION NUMBER: 60/080,114
;  PRIOR FILING DATE: 1998-03-31
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 23767
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5864
;   LENGTH: 808
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 16, 20, 57, 68, 77, 78, 80, 89, 105, 107, 108, 115, 117,
;   LOCATION: 120, 123, 129, 133, 138, 140, 143, 147, 148, 156, 159, 163,
;   LOCATION: 166, 172, 173, 174, 180, 186, 192, 197, 198, 201, 207, 209,
;   LOCATION: 214, 216, 217, 218, 219, 226, 228, 236, 237, 238, 243
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 244, 245, 246, 254, 283, 308, 311, 317, 325, 330, 332, 339,
;   LOCATION: 357, 358, 375, 385, 398, 401, 415, 422, 424, 426, 430, 434,
;   LOCATION: 442, 448, 449, 451, 470, 471, 475, 476, 477, 479, 480, 491,
;   LOCATION: 497, 507, 510, 527, 559, 574, 576, 584, 591, 609, 636
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 639, 641, 643, 650, 660, 672, 673, 681, 684, 722, 743, 754,
;   LOCATION: 762, 783, 784, 795, 801
;   OTHER INFORMATION: n = A,T,C or G
US-10-779-543-5864

  Query Match             45.8%;  Score 22;  DB 6;  Length 808;
  Best Local Similarity   68.3%;  Pred. No. 5.9;
  Matches   28;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGC 41
              ||||||   |||   ||| ||| |  ||||||||||   ||
Db        333 TATGATNTTCTGCCCACCTGTGATNNTACCCATGAACACGC 373


RESULT 7
US-10-767-471-34311
; Sequence 34311, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34311
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-34311

  Query Match             45.0%;  Score 21.6;  DB 6;  Length 201;
  Best Local Similarity   68.2%;  Pred. No. 6.4;
  Matches   30;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              | ||||||| |   ||  | ||| | |||  |||||| || |||
Db        155 TTTGATCAGTTTCATATTCCTGTGGGTACTAATGAAAAGGAACA 198


RESULT 8
US-10-767-701-14458
; Sequence 14458, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14458
;   LENGTH: 2895
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS8018_1
US-10-767-701-14458

  Query Match             45.0%;  Score 21.6;  DB 6;  Length 2895;
  Best Local Similarity   68.2%;  Pred. No. 12;
  Matches   30;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              || | || | |||| |   ||||||  | |||||| | || |||
Db       1033 ATAAGCAACGGGTTGCAGTTGTTGTAGCTCATGAATTAGCTCAT 1076


RESULT 9
US-10-767-471-10717
; Sequence 10717, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10717
;   LENGTH: 546025
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(546025)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-767-471-10717

  Query Match             45.0%;  Score 21.6;  DB 6;  Length 546025;
  Best Local Similarity   68.2%;  Pred. No. 36;
  Matches   30;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              | ||||||| |   ||  | ||| | |||  |||||| || |||
Db     329443 TTTGATCAGTTTCATATTCCTGTGGGTACTAATGAAAAGGAACA 329486


RESULT 10
US-10-767-701-31393/c
; Sequence 31393, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
```

; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 31393
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(607)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: 18067944
US-10-767-701-31393

Query Match 44.6%; Score 21.4; DB 6; Length 607;
Best Local Similarity 66.0%; Pred. No. 9.8;
Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
Db 306 AGGATGGGCTGGTGAGCTGTTTATGCACGCATTACATGGCACAAACA 260

RESULT 11
US-10-765-790-41/c
; Sequence 41, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 26371
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-41

Query Match 43.8%; Score 21; DB 6; Length 26371;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
Db 15911 AAGCCCAGCTGGTGACTTGAGTGTCTCCCCAACAAATAGAACATG 15867

RESULT 12
US-10-783-271-34
; Sequence 34, Application US/10783271
; GENERAL INFORMATION:
; APPLICANT: Veridex, LLC
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: BREAST CANCER PROGNOSTICS
; FILE REFERENCE: VDX-5003 USNP
; CURRENT APPLICATION NUMBER: US/10/783,271
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 175737
; TYPE: DNA
; ORGANISM: human
US-10-783-271-34

Query Match 43.3%; Score 20.8; DB 6; Length 175737;
Best Local Similarity 70.0%; Pred. No. 60;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 9 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
Db 116051 GCTGCTCATCTGGGACTTTGCCCCTGAAATGGCACCTCCA 116090

RESULT 13
US-10-779-543-15761/c
; Sequence 15761, Application US/10779543
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15761
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-779-543-15761

Query Match 42.9%; Score 20.6; DB 6; Length 421;
Best Local Similarity 74.3%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 11 TGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
Db 225 TTGTTAACTGTCCAGTTACCCATGAAAATTCACAT 191

RESULT 14
US-10-781-469-49
; Sequence 49, Application US/10781469
; GENERAL INFORMATION:
; APPLICANT: Reddy, Roopa
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice K.
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING RECEPTOR PROTEINS
; FILE REFERENCE: PA-0014-1 CON
; CURRENT APPLICATION NUMBER: US/10/781,469
; CURRENT FILING DATE: 2004-02-17
; PRIOR APPLICATION NUMBER: US 09/496,321
; PRIOR FILING DATE: 2000-02-01

```
;  PRIOR APPLICATION NUMBER: US 60/118,905
;  PRIOR FILING DATE: 1999-02-05
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: PERL Program
; SEQ ID NO 49
;   LENGTH: 905
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 2832568
US-10-781-469-49

  Query Match             42.9%;  Score 20.6;  DB 6;  Length 905;
  Best Local Similarity   67.4%;  Pred. No. 23;
  Matches   29;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              || | | |  ||| ||| | ||  |  |||||||| |||| ||
Db        605 ATAAACTGTGGGTCACCAGAGTCATAGCCCATGAACTGGCGCA 647


RESULT 15
US-10-775-169-52
; Sequence 52, Application US/10775169
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Burczynski, Michael
;  APPLICANT:  Twine, Natalie
;  APPLICANT:  Dorner, Andrew
;  APPLICANT:  Trepicchio, William
;  TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
;  FILE REFERENCE: AM101080 (031896-013000)
;  CURRENT APPLICATION NUMBER: US/10/775,169
;  CURRENT FILING DATE:  2004-02-11
;  NUMBER OF SEQ ID NOS: 5278
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 52
;   LENGTH: 198161
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-169-52

  Query Match             42.9%;  Score 20.6;  DB 6;  Length 198161;
  Best Local Similarity   67.4%;  Pred. No. 74;
  Matches   29;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||  ||  | | | |||| ||| ||| |||||| | || | |
Db     175379 TCACATGTGTTCTCATGTTCTTAACCACGAAATGTCTCAGGGA 175421


Search completed: March 19, 2004, 03:10:06
Job time : 8.69942 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 214.058 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25.2	52.5	878	28	CC001571	CC001571 PUDIE23TD
2	25	52.1	894	13	BU147350	BU147350 AGENCOURT
3	24.8	51.7	273	10	BB435111	BB435111 BB435111
c 4	24.6	51.3	1201	9	AL576676	AL576676 AL576676
5	24.2	50.4	446	9	AU299906	AU299906 AU299906
c 6	24.2	50.4	539	28	AZ056038	AZ056038 RPCI-23-4
7	24.2	50.4	903	14	CB200568	CB200568 AGENCOURT
c 8	24.2	50.4	1032	29	CNS01XVV	AL172228 Tetraodon
9	24	50.0	176	10	AW837303	AW837303 QV2-LT003
10	24	50.0	300	10	BF634308	BF634308 NF082C02D
11	24	50.0	600	12	BJ102563	BJ102563 BJ102563
c 12	24	50.0	610	9	AL589140	AL589140 AL589140
c 13	24	50.0	766	13	BU294276	BU294276 603606963
c 14	24	50.0	780	13	BU252275	BU252275 603748238
c 15	24	50.0	791	13	BU324056	BU324056 603493436
c 16	24	50.0	812	13	BU328355	BU328355 603406191
c 17	24	50.0	813	13	BU202575	BU202575 604152384
c 18	23.8	49.6	294	9	AA224363	AA224363 zr14b09.s
c 19	23.6	49.2	296	10	AW373707	AW373707 QV4-BT053
c 20	23.6	49.2	316	14	CB017485	CB017485 pgn1c.pk0
c 21	23.6	49.2	994	13	BU250654	BU250654 603321055
22	23.4	48.8	446	10	BE443880	BE443880 WHE1124_F
23	23.4	48.8	605	14	CB891544	CB891544 EST648513
c 24	23.4	48.8	673	14	CD452770	CD452770 WHE1124_F
25	23.4	48.8	1016	29	CC600714	CC600714 ZMMBBc040
c 26	23.2	48.3	153	13	BQ172708	BQ172708 1091027E1
27	23.2	48.3	230	10	BB435292	BB435292 BB435292
c 28	23.2	48.3	248	14	CF245156	CF245156 3530_1_8_
29	23.2	48.3	275	10	BB435052	BB435052 BB435052
30	23.2	48.3	295	10	BB436405	BB436405 BB436405
31	23.2	48.3	305	10	BB433896	BB433896 BB433896
c 32	23.2	48.3	460	12	BM379062	BM379062 MEST573-G
c 33	23.2	48.3	467	13	BQ744659	BQ744659 946106F04
34	23.2	48.3	479	12	BG947566	BG947566 IP1_6_B09
35	23.2	48.3	521	14	CD204019	CD204019 HS1_3_C07
36	23.2	48.3	522	12	BI076348	BI076348 IP1_26_B0
c 37	23.2	48.3	540	9	AI711846	AI711846 605064G07
c 38	23.2	48.3	545	10	BE358333	BE358333 DG1_29_D0
c 39	23.2	48.3	577	14	CA404413	CA404413 EL01N0517
40	23.2	48.3	590	12	BG103593	BG103593 RHIZ2_38_
41	23.2	48.3	591	12	BM318096	BM318096 PI1_78_D1
42	23.2	48.3	615	10	BE365335	BE365335 PI1_26_B1
c 43	23.2	48.3	620	14	CF049562	CF049562 QCL38b09.
44	23.2	48.3	623	10	BE919209	BE919209 FM1_4_E01
c 45	23.2	48.3	623	14	CD527595	CD527595 3529_1_12

ALIGNMENTS

RESULT 1
CC001571
LOCUS CC001571 878 bp DNA linear GSS 31-MAR-2003
DEFINITION PUDIE23TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa217C21, genomic survey sequence.
ACCESSION CC001571
VERSION CC001571.1 GI:29380131
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 878)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.

```
FEATURES             Location/Qualifiers
     source          1. .878
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa217C21"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             52.5%;  Score 25.2;  DB 28;  Length 878;
  Best Local Similarity   71.7%;  Pred. No. 87;
  Matches   33;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              |||| ||||  |||  |||  |||  | ||||| |||| || ||||
Db        548 ATGACCAGCAAGTTGTCCGACTTGAAAACCATGGAATGACAGATGC 593


RESULT 2
BU147350
LOCUS       BU147350                894 bp    mRNA    linear   EST 03-SEP-2002
DEFINITION  AGENCOURT_8070809 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6089930
            5', mRNA sequence.
ACCESSION   BU147350
VERSION     BU147350.1  GI:22660882
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2330  row: i  column: 03
            High quality sequence stop: 713.
FEATURES             Location/Qualifiers
     source          1. .894
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6089930"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_112"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             52.1%;  Score 25;  DB 13;  Length 894;
  Best Local Similarity   75.6%;  Pred. No. 1e+02;
  Matches   31;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              |||| ||||   ||||| ||||||||||| |  ||  ||||


Db        841 ATGAGCAGCCCCTTACCAGTGTTGTTACCAACCAAGGGGCA 881


RESULT 3
BB435111
LOCUS       BB435111                273 bp    mRNA    linear   EST 01-AUG-2000
DEFINITION  BB435111 RIKEN full-length enriched, adult pancreas Islet cells Mus
            musculus cDNA clone C820011N02 3', mRNA sequence.
ACCESSION   BB435111
VERSION     BB435111.1  GI:9274838
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
            Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
            Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
            Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
            Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A.,
            Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
            Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A.,
            Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .273
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="C820011N02"
                     /tissue_type="pancreas"
                     /cell_type="Islet cells"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult pancreas
                     Islet cells"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
```

primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I.~Islet cells were provided by Hiroo Iwata, Institute for Frontier Medical Sciences, Kyoto University, Sakyo-ku, Kyoto, 606-8507 Japan, whose assistance we gratefully acknowledge."

ORIGIN

Query Match 51.7%; Score 24.8; DB 10; Length 273;
Best Local Similarity 72.7%; Pred. No. 84;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
Db 147 ATGACCACCTGGTTACCCTAGTGGTTAACCTGGAAAGAGAACTT 190

RESULT 4
AL576676/c
LOCUS AL576676 1201 bp mRNA linear EST 01-JUN-2003
DEFINITION AL576676 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI080YF21 3-PRIME, mRNA sequence.
ACCESSION AL576676
VERSION AL576676.2 GI:31314957
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 16, 2001 this sequence version replaced gi:12939053.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 1179.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DI080CC11NP1&cluster=1179.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DI080CC11NP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI080YF21"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 51.3%; Score 24.6; DB 9; Length 1201;
Best Local Similarity 53.3%; Pred. No. 1.6e+02;
Matches 24; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
Db 1093 WMYGRGMCGCRGKATACCCYKTAATTTACMCATGAAATGRSCCAW 1049

RESULT 5
AU299906
LOCUS AU299906 446 bp mRNA linear EST 17-JUL-2003
DEFINITION AU299906 Cryptomeria japonica developing female cone Cryptomeria japonica cDNA clone CF0701 5', mRNA sequence.
ACCESSION AU299906
VERSION AU299906.1 GI:32947357
KEYWORDS EST.
SOURCE Cryptomeria japonica (Japanese cedar)
ORGANISM Cryptomeria japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.
REFERENCE 1 (bases 1 to 446)
AUTHORS Ujino-Ihara,T., Yoshimura,K. and Tsumura,Y.
TITLE Expressed sequences in developing flower tissues of Cryptomeria japonica
JOURNAL Unpublished (2003)
COMMENT Contact: Tokuko Ujino-Ihara
Bio-resources Technology Division
Forestry and Forest Products Research Institute
Matsunosato 1, Kukizaki, Ibaraki 305-8687, Japan
Tel: 81-298-73-3211(ex.444)
Fax: 81-298-73-3795
Email: udino@ffpri.affrc.go.jp;
URL:http://www.ffpri.affrc.go.jp/labs/cjgenome/database/cjdatae.htm1.
FEATURES Location/Qualifiers
source 1. .446
/organism="Cryptomeria japonica"
/mol_type="mRNA"
/db_xref="taxon:3369"
/clone="CF0701"
/tissue_type="developing female cone"
/clone_lib="Cryptomeria japonica developing female cone"

ORIGIN

Query Match 50.4%; Score 24.2; DB 9; Length 446;
Best Local Similarity 71.1%; Pred. No. 1.7e+02;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
Db 141 GATCACCTTGATAACCCTGTTGCTAACAATGCATGGGAATATGCA 185

RESULT 6
AZ056038/c
LOCUS AZ056038 539 bp DNA linear GSS 30-MAR-2000
DEFINITION RPCI-23-413M16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-413M16, genomic survey sequence.
ACCESSION AZ056038
VERSION AZ056038.1 GI:7347274
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 539)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-413M16.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

```
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 413  row: M  column: 16
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-413M16"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
ORIGIN

  Query Match             50.4%;  Score 24.2;  DB 28;  Length 539;
  Best Local Similarity   71.1%;  Pred. No. 1.8e+02;
  Matches   32;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              | |||| || || |   |||||| ||||||  |||| ||||  ||
Db        493 GTTCAGATGTTTGCAATTGTTGTGACCCATTCAATGTCACAGTCA 449


RESULT 7
CB200568
LOCUS       CB200568                903 bp    mRNA    linear   EST 21-MAR-2003
DEFINITION  AGENCOURT_11271837 NICHD_XGC_Tad2 Xenopus laevis cDNA clone
            IMAGE:6875681 5', mRNA sequence.
ACCESSION   CB200568
VERSION     CB200568.2  GI:29147154
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Feb 4, 2003 this sequence version replaced gi:28233219.
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Drs. Donald Brown and Liquan Cai
             cDNA Library Preparation: CLONTECH
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3111  row: h  column: 16
            High quality sequence stop: 104.
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6875681"
                     /dev_stage="metamorphosis stage 62"
                     /clone_lib="NICHD_XGC_Tad2"
                     /note="Organ: Developing Tadpole; Vector: pDNR-LIB;
                     Site_1: SfiI; Site_2: SfiI; 5' and 3' adaptors were used
                     in cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.7 kb
                     (range 0.8-3.0 kb). 15/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA)."
ORIGIN

  Query Match             50.4%;  Score 24.2;  DB 14;  Length 903;
  Best Local Similarity   71.1%;  Pred. No. 2.1e+02;
  Matches   32;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              |||||  ||||   |||  | |  ||||   |||||||  |  ||||||
Db        434 ATGATGAGCTCCCTACTCTTATTGTGCTCCATGAAGTCTCACATG 478


RESULT 8
CNS01XVV/c
LOCUS       CNS01XVV                1032 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            217C04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL172228
VERSION     AL172228.1  GI:7810285
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1032)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="217C04"
```

```
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG217BB02SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             50.4%;  Score 24.2;  DB 29;  Length 1032;
  Best Local Similarity   71.1%;  Pred. No. 2.2e+02;
  Matches   32;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              | |  |||||||||| |.| |||||||| || |  | || || ||
Db        573 AAGCACAGCTGGTTATCAGAGTTGTTACACACGTGAAGGAACCTG 529


RESULT 9
AW837303
LOCUS       AW837303                 176 bp    mRNA    linear   EST 18-MAY-2000
DEFINITION  QV2-LT0038-180300-079-d01 LT0038 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW837303
VERSION     AW837303.1  GI:7931277
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV2-LT0038-180
            300-079-d01&t3=2000-03-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 26
            High quality sequence stop: 176.
FEATURES             Location/Qualifiers
     source          1. .176
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="LT0038"
                     /note="Organ: leiomios; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             50.0%;  Score 24;  DB 10;  Length 176;
  Best Local Similarity   68.8%;  Pred. No. 1.4e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  ||||||| | ||||| ||  || ||| | ||||    ||||||
Db          4 TACAATCAGCTCGGTACCCAGGTCTTTTCCCCTCAAATCATGCATGCA 51


RESULT 10
BF634308
LOCUS       BF634308                 300 bp    mRNA    linear   EST 19-DEC-2000
DEFINITION  NF082C02DT1F1017 Drought Medicago truncatula cDNA clone NF082C02DT
            5', mRNA sequence.
ACCESSION   BF634308
VERSION     BF634308.1  GI:11898466
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula drought library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7391
            Fax: 580 221 7380
            Email: gdmay@noble.org
            Insert Length: 300   Std Error: 0.00
            Plate: 082  row: C  column: 02
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF082C02DT"
                     /tissue_type="Plantlets"
                     /dev_stage="Pooled timepoints"
                     /clone_lib="Drought"
                     /note="Vector: Lambda Zap; Contains a mixture of entire
                     plantlets harvested in a series of days-post-watering
                     timepoints. "
ORIGIN

  Query Match             50.0%;  Score 24;  DB 10;  Length 300;
  Best Local Similarity   73.2%;  Pred. No. 1.7e+02;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          5 ATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | |||  |||| |   || ||| |||||||||| | ||||||
Db        221 AACAGAGGGTTGCTGTTGNTGTAACCCATGAACTAGCACAT 261


RESULT 11
BJ102563
LOCUS       BJ102563                 600 bp    mRNA    linear   EST 30-MAY-2003
DEFINITION  BJ102563 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1040a11 5', mRNA sequence.
ACCESSION   BJ102563
VERSION     BJ102563.2  GI:31238878
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
```

```
  JOURNAL   Unpublished (2002)
COMMENT     On Jan 18, 2002 this sequence version replaced gi:18245233.
            Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Caenorhabditis elegans"
                     /mol_type="mRNA"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1040a11"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"
ORIGIN

  Query Match             50.0%;  Score 24;  DB 12;  Length 600;
  Best Local Similarity   68.8%;  Pred. No. 2.2e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||| ||| | | ||| || | || ||| |||| ||| | ||  ||||
Db        299 TATATTCACCGGATTATCCATATTATTATCCATCAAAAGTCAATTGCA 346


RESULT 12
AL589140/c
LOCUS       AL589140                610 bp    mRNA    linear   EST 02-MAR-2001
DEFINITION  AL589140 BP Chicken Brain Library Gallus gallus cDNA clone
            ROS083H01, mRNA sequence.
ACCESSION   AL589140
VERSION     AL589140.1  GI:13194174
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Murray,F.
  TITLE     BP Chicken Brain Library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Frazer Murray
            Dept. Genomics and Bioinformatics
            Roslin Institute
            Roslin, Midlothian, EH25 9PS, UK
            Tel: +44 (0)131 527 4200
            Fax: +44 (0)131 440 0434
            Email: frazer.murray@bbsrc.ac.uk
            GCGGCCGCTTTTTTTTTTTTTTTTT 3' Poly A RNA purchased from Clonetech
            (*6854-
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9031"
                     /clone="ROS083H01"
                     /tissue_type="Brain"
                     /dev_stage="Unknown"
                     /lab_host="DH10B"
                     /clone_lib="BP Chicken Brain Library"
                     /note="Vector: pSPORT1; Site_1: Not1; Site_2: Sal1; Cloned
                     unidirectionally. Primer: Oligo dT. 5' adaptor sequence:
                     5' TCGACCTCGAG 3' ; 3' adaptor sequence: 5'
                     GCGGCCGCTTTTTTTTTTTTTTTTT 3' Poly A RNA purchased from
                     Clonetech (*6854-1)"
ORIGIN

  Query Match             50.0%;  Score 24;  DB 9;  Length 610;
  Best Local Similarity   68.8%;  Pred. No. 2.2e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  || | ||||||| | |||||||||  |  | |||| || ||  ||
Db        397 TACAATGATCTGGTTATCTGTGTTGTTTACAGTAAAATAGCCCACACA 350


RESULT 13
BU294276/c
LOCUS       BU294276                766 bp    mRNA    linear   EST 27-NOV-2002
DEFINITION  603606963F1 CSEQCHN55 Gallus gallus cDNA clone ChEST590n16 5', mRNA
            sequence.
ACCESSION   BU294276
VERSION     BU294276.1  GI:25743899
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 766)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk
FEATURES             Location/Qualifiers
     source          1. .766
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 15I"
                     /db_xref="taxon:9031"
                     /clone="ChEST590n16"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN55"
                     /note="Organ: kidney + adrenal; Vector: pBluescript II
                     KS(+); Site_1: EcoRI; Site_2: Not1; This normalized
                     library was constructed from 1 million independent clones.
                     cDNA synthesis was initiated using an oligo(dT) primer,
                     using methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

  Query Match             50.0%;  Score 24;  DB 13;  Length 766;
  Best Local Similarity   68.8%;  Pred. No. 2.4e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  || | ||||||| | |||||||||  |  | |||| || ||  ||
```

```
Db        183 TACAATGATCTGGTTATCTGTGTTGTTTACAGTAAAATAGCCCACACA 136


RESULT 14
BU252275/c
LOCUS       BU252275                780 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  603748238F1 CSEQCHN38 Gallus gallus cDNA clone ChEST656l10 5', mRNA
            sequence.
ACCESSION   BU252275
VERSION     BU252275.1  GI:25505294
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 780)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .780
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST656l10"
                     /dev_stage="36"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN38"
                     /note="Organ: limbs; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

  Query Match             50.0%;  Score 24;  DB 13;  Length 780;
  Best Local Similarity   68.8%;  Pred. No. 2.4e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  || | ||||||| | ||||||||  |  | |||| || ||  ||
Db        331 TACAATGATCTGGTTATCTGTGTTGTTTACAGTAAAATAGCCCACACA 284


RESULT 15
BU324056/c
LOCUS       BU324056                791 bp    mRNA    linear   EST 28-NOV-2002
DEFINITION  603493436F1 CSEQCHN63 Gallus gallus cDNA clone ChEST399g10 5', mRNA
            sequence.
ACCESSION   BU324056
VERSION     BU324056.1  GI:25832057
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST399g10"
                     /dev_stage="36"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN63"
                     /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

  Query Match             50.0%;  Score 24;  DB 13;  Length 791;
  Best Local Similarity   68.8%;  Pred. No. 2.4e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  || | ||||||| | ||||||||  |  | |||| || ||  ||
Db        369 TACAATGATCTGGTTATCTGTGTTGTTTACAGTAAAATAGCCCACACA 322


Search completed: March 18, 2004, 20:11:24
Job time : 218.058 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		456	100.0	456	6	I44909	I44909 Sequence 58
2		456	100.0	2067	1	SEU43280	U43280 Salmonella
3		448	98.2	5103	1	STAJ2301	AJ002301 Salmonell
4		448	98.2	22411	1	AE008749	AE008749 Salmonell
5		446.4	97.9	254050	1	AL627269	AL627269 Salmonell
6	c	446.4	97.9	301983	1	AE016840	AE016840 Salmonell
7		440.6	96.6	1048	1	STAGFBA	AJ000514 Salmonell
8		331.8	72.8	361	6	I44908	I44908 Sequence 56
9		310.4	68.1	2889	1	CSP515700	AJ515700 Citrobact
10		262.8	57.6	2920	1	CFR515701	AJ515701 Citrobact
11		256	56.1	4680	1	ECCSGABDG	X90754 E.coli csgG
12		256	56.1	10346	1	AE000205	AE000205 Escherich
13		256	56.1	15047	1	D90741	D90741 Escherichia
14		254.4	55.8	456	6	AX814811	AX814811 Sequence
15		254.4	55.8	648	1	ECOCSGAA	L04979 Escherichia
16		247.8	54.3	1711	1	AF275733	AF275733 Escherich
17		247.8	54.3	10190	1	AE005315	AE005315 Escherich
18		247.8	54.3	306358	1	AE016759	AE016759 Escherich
19		247.8	54.3	327773	1	AP002554	AP002554 Escherich
20		229	50.2	230	1	SEU53207	U53207 Salmonella
21		217.6	47.7	2883	1	ESA515702	AJ515702 Enterobac
22		190.4	41.8	10370	1	AE015131	AE015131 Shigella
23		190.4	41.8	292504	1	AE016981	AE016981 Shigella
24	c	167.4	36.7	437	1	AF237726	AF237726 Shigella
25		93.2	20.4	19201	1	D90742	D90742 Escherichia
26		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27		48.2	10.6	78	6	AX814809	AX814809 Sequence
28		44.8	9.8	72	6	AX814798	AX814798 Sequence
29	c	44	9.6	2000	6	AX655393	AX655393 Sequence
30	c	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
31	c	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
32	c	36	7.9	168267	2	AC146009	AC146009 Pan trogl
33		35.6	7.8	301214	1	AE016786	AE016786 Pseudomon
34		34.6	7.6	11160	1	AE008209	AE008209 Agrobacte
35	c	34.6	7.6	11476	1	AE009405	AE009405 Agrobacte
36		34.4	7.5	151828	2	AC146008	AC146008 Pan trogl
37		34.2	7.5	1458	1	AF173043	AF173043 Orientia
38		34	7.5	956	8	AK058943	AK058943 Oryza sat
39	c	34	7.5	1344	6	BD243128	BD243128 Hypersens
40	c	34	7.5	1344	6	BD243666	BD243666 Hypersens
41	c	34	7.5	1344	6	AR403552	AR403552 Sequence
42	c	34	7.5	1344	6	AX287002	AX287002 Sequence
43	c	34	7.5	1344	6	AX343147	AX343147 Sequence
44	c	34	7.5	1344	6	AX370155	AX370155 Sequence
45	c	34	7.5	1344	6	AX536272	AX536272 Sequence

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 456;  DB 6;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-120;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match             100.0%;  Score 456;  DB 1;  Length 2067;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-120;
```

```
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIPVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match             98.2%;  Score 448;  DB 1;  Length 5103;
  Best Local Similarity   98.9%;  Pred. No. 3.9e-118;
  Matches  451;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA

COMMENT COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES Location/Qualifiers

source 1. .22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

gene 434. .1308
/gene="phoH"
/note="synonym: STM1126"

RBS 434. .439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"

CDS 454. .1308
/gene="phoH"
/note="similar to E. coli PhoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=11
/product="PhoB-dependent ATP-binding pho regulon component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
QRTLHAYS"

gene complement(1414. .2303)
/gene="STM1127"

CDS complement(1414. .2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29% identity in aa 15 - 264"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
QARESKLRTMNALDMTK"

RBS complement(2298. .2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"

gene complement(2580. .4088)
/gene="STM1128"

CDS complement(2580. .4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22% identity in aa 7 - 478"
/codon_start=1
/transl_table=11
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"

RBS complement(4083. .4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"

gene complement(4413. .5099)
/gene="STM1129"

CDS complement(4413. .5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 - 226"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
WFSHAVKR"

RBS complement(5094. .5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"

gene 5599. .6759
/gene="STM1130"

CDS 5599. .6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41% identity in aa 40 - 403"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"

gene 6793. .7497
/gene="STM1131"

RBS 6793. .6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           98.2%;  Score 448;  DB 1;  Length 22411;
  Best Local Similarity   98.9%;  Pred. No. 4.3e-118;
  Matches  451;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||| |||||||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSPEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             97.9%;  Score 446.4;  DB 1;  Length 254050;
  Best Local Similarity   98.7%;  Pred. No. 1.5e-117;
  Matches  450;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| ||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840              301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
gene            complement(1812. .3473)
                /gene="flgK"
                /locus_tag="t1737"
CDS             complement(1812. .3473)
                /gene="flgK"
                /locus_tag="t1737"
                /note="corresponds to STY1222 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook-associated protein 1"
                /protein_id="AAO69361.1"
                /db_xref="GI:29137800"
                /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
gene            complement(3538. .4488)
                /gene="flgJ"
                /locus_tag="t1738"
CDS             complement(3538. .4488)
                /gene="flgJ"
                /locus_tag="t1738"
                /note="corresponds to STY1221 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar protein FlgJ"
                /protein_id="AAO69362.1"
                /db_xref="GI:29137801"
                /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
gene            complement(4488. .5591)
                /gene="flgI"
                /locus_tag="t1739"
CDS             complement(4488. .5591)
                /gene="flgI"
                /locus_tag="t1739"
                /note="corresponds to STY1220 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar P-ring protein precursor"
                /protein_id="AAO69363.1"
                /db_xref="GI:29137802"
                /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             97.9%;  Score 446.4;  DB 1;  Length 301983;
Best Local Similarity   98.7%;  Pred. No. 1.5e-117;
  Matches  450;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             96.6%;  Score 440.6;  DB 1;  Length 1048;
  Best Local Similarity   98.0%;  Pred. No. 4.9e-116;
  Matches  446;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  ||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             72.8%;  Score 331.8;  DB 6;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 1.2e-84;
  Matches  333;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           68.1%;  Score 310.4;  DB 1;  Length 2889;
  Best Local Similarity  81.6%;  Pred. No. 2.1e-78;
  Matches  372;  Conservative    0;  Mismatches   81;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| ||||||||||||||||   |    |||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  ||||||||||||||  ||   |  |||||| |||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||    ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || ||||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || |||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
  Query Match             57.6%;  Score 262.8;  DB 1;  Length 2920;
  Best Local Similarity   75.2%;  Pred. No. 1.2e-64;
  Matches  343;  Conservative    0;  Mismatches  107;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||    ||| ||||   ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||    ||||| || ||   |||  ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||| ||| | ||||| |||   |||| ||||  || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| ||||| |||||||| || |||||  | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||| |||||||||||||| || ||||| ||||  ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

```
RESULT 11
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
```

```
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729...3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           56.1%;  Score 256;  DB 1;  Length 4680;
  Best Local Similarity  72.6%;  Pred. No. 1.1e-62;
  Matches  331;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||| || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy         301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
               || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy         361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               ||||  ||    ||| | ||||| |||||| | |||  ||| | |||  |||||||
Db        4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy         421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
               || ||||| |||||||| || ||| | ||||| |||
Db        4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 12
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
```

```
          and products; all new functional assignments courtesy of Monica
          Riley; added promoters, protein binding sites, and repeated
          sequences described in reference 1. The unique numeric identifiers
          beginning with a lowercase 'b' assigned to each gene (protein- or
          RNA-encoding) are now designated as gene synonyms instead of
          labels. This should allow them to be searched for in Entrez as gene
          names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

Query Match             56.1%;  Score 256;  DB 1;  Length 10346;
Best Local Similarity   72.6%;  Pred. No. 1.2e-62;
Matches  331;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||    | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||| ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

RESULT 13
D90741
LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
FEATURES Location/Qualifiers
source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER

```
                ILNVSPRRLLNFLESRGMAPIAEFADL"
     CDS        2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
     CDS        3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
     gene       complement(4117. .4950)
                /gene="csgG"
     CDS        complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
     gene       complement(4977. .5393)
                /gene="csgF"
     CDS        complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene       complement(5418. .5807)
                /gene="csgE"
     CDS        complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
     gene       complement(5812. .6462)
                /gene="csgD"
     CDS        complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene       7217. .7672
                /gene="csgB"
     CDS        7217. .7672

  Query Match          56.1%;  Score 256;  DB 1;  Length 15047;
  Best Local Similarity  72.6%;  Pred. No. 1.2e-62;
  Matches  331;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || |||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||     |||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 14
AX814811
LOCUS       AX814811                 456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
```

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match             55.8%;  Score 254.4;  DB 6;  Length 456;
  Best Local Similarity   72.4%;  Pred. No. 2.8e-62;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match             55.8%;  Score 254.4;  DB 1;  Length 648;
  Best Local Similarity   72.4%;  Pred. No. 2.9e-62;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 18, 2004, 15:02:22
Job time : 2314.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	2	AAQ87467	Aaq87467 AgfA sequ
	2	456	100.0	456	2	AAT74142	Aat74142 Salmonell
	3	456	100.0	456	3	AAC64617	Aac64617 Salmonell
	4	412.8	90.5	456	3	AAC64626	Aac64626 AgfA::PT3
	5	406.4	89.1	456	3	AAC64625	Aac64625 AgfA::PT3
	6	404.8	88.8	456	3	AAC64628	Aac64628 AgfA::PT3
	7	404.8	88.8	456	3	AAC64622	Aac64622 AgfA::PT3
	8	404.8	88.8	456	3	AAC64629	Aac64629 AgfA::PT3
	9	404.8	88.8	456	3	AAC64623	Aac64623 AgfA::PT3
	10	403.2	88.4	456	3	AAC64624	Aac64624 AgfA::PT3
	11	401.6	88.1	456	3	AAC64630	Aac64630 AgfA::PT3
	12	401.6	88.1	456	3	AAC64627	Aac64627 AgfA::PT3
	13	398.4	87.4	456	3	AAC64631	Aac64631 AgfA::PT3
	14	331.8	72.8	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	331.8	72.8	361	2	AAT74141	Aat74141 Salmonell
	16	254.4	55.8	456	3	AAC64619	Aac64619 Escherich
	17	254.4	55.8	456	9	ACF36153	Acf36153 E. coli C
	18	242	53.1	646	2	AAQ62647	Aaq62647 Fibronect
	19	180	39.5	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
	21	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	22	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
c	23	44	9.6	2000	7	ADA71938	Ada71938 Rice gene
	24	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
	25	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
c	26	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	27	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
	28	34.6	7.6	456	3	AAC64620	Aac64620 Escherich
	29	34	7.5	456	3	AAC64618	Aac64618 Salmonell
c	30	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	31	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	32	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	33	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	34	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	35	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c	36	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a
	37	34	7.5	2886	7	ACC59886	Acc59886 Saccharom
	38	34	7.5	3300	3	AAZ55699	Aaz55699 DNA encod
	39	33.4	7.3	3990	2	AAX61785	Aax61785 B. burgdo
	40	33.4	7.3	5100	2	AAX20271	Aax20271 Borrelia
	41	33.4	7.3	7766	2	AAT85043	Aat85043 Borrelia
	42	33.2	7.3	788	4	AAH08100	Aah08100 Human cDN
c	43	32.8	7.2	78	3	AAC64609	Aac64609 AgfA (SEF
	44	32.8	7.2	1005	7	ACA02133	Aca02133 C. glutam
	45	32.8	7.2	1059	5	AAH65447	Aah65447 C glutami

ALIGNMENTS

RESULT 1
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX
FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)

```
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 456;  DB 2;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 8.2e-138;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
AAT74142
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 456;  DB 2;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 8.2e-138;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
```

PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36341.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.2e-138;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 4
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 3; Length 456;
Best Local Similarity 94.1%; Pred. No. 9.2e-124;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||| |||   |   ||||||   | |     | |   ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 3; Length 456;
Best Local Similarity 93.2%; Pred. No. 1.1e-121;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |    ||  |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
AAC64628
ID AAC64628 standard; DNA; 456 BP.
XX
AC AAC64628;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36352.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match   88.8%; Score 404.8; DB 3; Length 456;
  Best Local Similarity 93.0%; Pred. No. 3.7e-121;
  Matches 424; Conservative   0; Mismatches   32; Indels   0; Gaps   0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||       ||   ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||  |    ||  |  |  ||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 7
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 3;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 3.7e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||| |           | | || | ||  |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 3;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 3.7e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||   | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 3;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 3.7e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64624
ID   AAC64624 standard; DNA; 456 BP.
XX
AC   AAC64624;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 88.4%; Score 403.2; DB 3; Length 456;
Best Local Similarity 92.8%; Pred. No. 1.2e-120;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 3; Length 456;
Best Local Similarity 92.5%; Pred. No. 4.1e-120;
Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |||  |     |  |  ||  |    |   |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36351.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 3; Length 456;
Best Local Similarity 92.5%; Pred. No. 4.1e-120;
Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |  |      || |  |    || |  |   | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             87.4%;  Score 398.4;  DB 3;  Length 456;
  Best Local Similarity   92.1%;  Pred. No. 4.5e-119;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||   || |||  |    |                |    | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             72.8%;  Score 331.8;  DB 2;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-97;
  Matches  333;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             72.8%;  Score 331.8;  DB 2;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-97;
  Matches  333;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 18, 2004, 09:11:45
```

Job time : 291.405 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	331.8	72.8	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	33.4	7.3	7766	4	US-09-125-619-3	Sequence 3, Appli
c	6	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	7	31.2	6.8	2436	4	US-09-540-236-492	Sequence 492, App
c	8	31.2	6.8	65792	4	US-09-596-002-31	Sequence 31, Appl
	9	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
	10	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	11	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	12	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	13	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	14	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	15	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
c	16	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
c	17	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	18	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	19	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	20	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	21	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	22	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, App
c	23	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, App
c	24	29.4	6.4	3494	4	US-09-023-655-1006	Sequence 1006, Ap
c	25	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
	26	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
c	27	29.2	6.4	2052	1	US-08-443-104-5	Sequence 5, Appli
c	28	29.2	6.4	2052	1	US-08-238-130-6	Sequence 6, Appli
c	29	29.2	6.4	2052	1	US-08-442-859-5	Sequence 5, Appli
c	30	29.2	6.4	2052	2	US-08-398-489-5	Sequence 5, Appli
c	31	29.2	6.4	2052	2	US-08-894-772-1	Sequence 1, Appli
c	32	29.2	6.4	2052	2	US-09-207-844-1	Sequence 1, Appli
c	33	29.2	6.4	2052	4	US-09-252-509-1	Sequence 1, Appli
c	34	29.2	6.4	2052	5	PCT-US95-05534-5	Sequence 5, Appli
	35	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	36	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	37	29	6.4	534	4	US-09-489-039A-2195	Sequence 2195, Ap
	38	29	6.4	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
c	39	29	6.4	13794	4	US-08-956-171E-54	Sequence 54, Appl
c	40	28.8	6.3	654	4	US-09-252-991A-4506	Sequence 4506, Ap
	41	28.8	6.3	720	4	US-09-328-352-354	Sequence 354, App
	42	28.8	6.3	1662	1	US-08-565-386-2	Sequence 2, Appli
	43	28.8	6.3	4529	1	US-08-565-386-1	Sequence 1, Appli
c	44	28.8	6.3	8906	4	US-09-027-169-5	Sequence 5, Appli
	45	28.4	6.2	1026	3	US-07-751-891B-24	Sequence 24, Appl

ALIGNMENTS

```
RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-788A-58

  Query Match           100.0%;  Score 456;  DB 1;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 3e-149;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 72.8%; Score 331.8; DB 1; Length 361;
Best Local Similarity 99.4%; Pred. No. 6e-106;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA AND ITS USE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,927
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,108
; FILING DATE: 06-AUG-1977
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1581
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304

```
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1344 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.12;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || |||| || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.12;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || |||| || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-09-125-619-3

  Query Match             7.3%;  Score 33.4;  DB 4;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 0.51;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||      ||  ||| |  | | |  | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |   |  | ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |  || ||| |   |   |  ||| |          | |||  |||||  |    |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||        | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 6
5486473-3/c
;Patent No. 5486473
;    APPLICANT: FUKITA, HIROYUKI;YOSHIDA, IWAO;TAKAGI, MITSUO;
;MANABE, SADAO;FUKAI, KONOSUKE
;    TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
;    NUMBER OF SEQUENCES: 16
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,049
;      FILING DATE: 09-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 809,255
;      FILING DATE: 18-DEC-1991
;      APPLICATION NUMBER: 279,685
;      FILING DATE: 05-DEC-1988
;      APPLICATION NUMBER: 932,419
;      FILING DATE: 19-NOV-1986
;SEQ ID NO:3:
;      LENGTH: 1491
5486473-3

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 1491;
  Best Local Similarity   50.0%;  Pred. No. 1.2;
  Matches   78;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||  |  || ||| |  |  | |||    | | | || | | |  | | |  | || |
Db        204 CAAGTCACTGACCGAAGCTAGGTAACAGTAACTGCGCACATCTGCGAGGTTGGCTGCTTC 145

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||    ||   ||     ||| ||    |||   |    ||  |||||  || || |
Db        144 CATGTTCATCATTTTGACATCAATGGTTGGCTTGTCTTTTGACATTATGGTCACACAACT 85
```

```
Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
              |  |||  |    |||||   ||| |    | ||||
Db         84 ATCGCCTTCCAGTACCAGATCAACCCATGTAGCTCC 49


RESULT 7
US-09-540-236-492/c
; Sequence 492, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2005-001
;  CURRENT APPLICATION NUMBER: US/09/540,236
;  CURRENT FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 492
;   LENGTH: 2436
;   TYPE: DNA
;   ORGANISM: M.catarrhalis
US-09-540-236-492

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 2436;
  Best Local Similarity   53.7%;  Pred. No. 1.6;
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |||||||| |    | ||  |       | | |  | || || ||||| |||   | |||
Db       1511 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 1452

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |     | |  || ||||  |||| |||||| |||| |||| ||  || |     |
Db       1451 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 1392

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |   || ||  ||  |||| ||  ||||   |  |||
Db       1391 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 1348


RESULT 8
US-09-596-002-31/c
; Sequence 31, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:      Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:      PERL Program
; SEQ ID NO 31
;   LENGTH: 65792
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte template ID No. 6632636 31
;  PUBLICATION INFORMATION:
US-09-596-002-31

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 65792;
  Best Local Similarity   53.7%;  Pred. No. 9.7;
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |||||||| |    | ||  |       | | |  | || || ||||| |||   | |||
Db       6819 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 6760

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |     | |  || ||||  |||| |||||| |||| |||| ||  || |     |
Db       6759 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 6700

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |   || ||  ||  |||| ||  ||||   |  |||
Db       6699 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 6656


RESULT 9
US-08-508-761B-1
; Sequence 1, Application US/08508761B
; Patent No. 6027920
;  GENERAL INFORMATION:
;    APPLICANT:  Joliff, Gwennael
;    APPLICANT:  Guyonvarch, Armel
;    APPLICANT:  Purification, Relano
;    APPLICANT:  Duchiron, Francis
;    APPLICANT:  Renaud, Michel
;    TITLE OF INVENTION:  System for Protein Expression and
;    TITLE OF INVENTION:  Secretion Especially in Corynebacteria
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jacobson, Price, Holman & Stern, PLLC
;      STREET:  400 Seventh St. N.W.
;      CITY:  Washington D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/508,761B
;      FILING DATE:  31-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09652
;      FILING DATE:  29-JUL-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09870
;      FILING DATE:  02-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Player, William E.
;      REGISTRATION NUMBER:  31,409
;      REFERENCE/DOCKET NUMBER:  P58525NA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 638-6666
;      TELEFAX:  (202) 393-5350
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2547 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Corynebacterium Melassecola
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  420..2390
US-08-508-761B-1

  Query Match             6.8%;  Score 30.8;  DB 3;  Length 2547;
  Best Local Similarity   58.9%;  Pred. No. 2.2;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
              |||| |||||  |  | | |  ||| ||| || ||||   ||||  | ||   ||  |||
```

```
Db        876 CAGAGTGGTTGGACCATTGAGACCAACATTGAGCAGTACTACGCCGATAAGAACGCCATT 935

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |||   || |  ||| ||  | | | ||
Db        936 GTTGTGCTCCCAATCGGTGGCGAGAGCTCC 965

RESULT 10
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             6.8%;  Score 30.8;  DB 4;  Length 1664976;
  Best Local Similarity   51.4%;  Pred. No. 71;
  Matches   71;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        288 TTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGG 347
              ||| | ||||  || ||  || | | |    ||||  || ||||    ||||| | ||
Db     360315 TTTTATAAATGTTGTCATGATGGTCAAAATCAACGAAAATAACTGTTTTATTATTCTCAT 360374

Qy        348 CCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAAT 407
              | |     |     || |||   ||  |  | ||  || | | | | |   ||    ||
Db     360375 CAACGGTGAAAACAAGAACAAAGCTTTTATCGATATGAACTCTTTTAAAATCATTTAAAG 360434

Qy        408 GGTGCGTCAGGTTGGTTT 425
              | ||  | |||||  | |
Db     360435 GATGTCTTAGGTTTTTAT 360452


RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4403765;
  Best Local Similarity   51.9%;  Pred. No. 90;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || ||||||| || ||  |  ||   | |||||| |||| |  |   || ||
Db    1357218 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357159

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
                ||     | | |    ||  | ||| | ||   |   ||   ||  || | |  |||
Db    1357158 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357099

Qy        165 TGCGCTTGCTCTG 177
                ||   |  |||
Db    1357098 CCCGGACGTGCTG 1357086


RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4411529;
  Best Local Similarity   51.9%;  Pred. No. 90;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || ||||||| || ||  |  ||   | |||||| |||| |  |   || ||
Db    1357750 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357691

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
                ||     | | |    ||  | ||| | ||   |   ||   ||  || | |  |||
Db    1357690 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357631

Qy        165 TGCGCTTGCTCTG 177
                ||   |  |||
Db    1357630 CCCGGACGTGCTG 1357618


RESULT 13
US-09-252-991A-10904
; Sequence 10904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10904
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904

Query Match 6.5%; Score 29.8; DB 4; Length 1008;
Best Local Similarity 54.0%; Pred. No. 3;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy          41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
               | | || | |  |  || | |  | ||| |  |    ||||  ||||||  | |     |
Db         141 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 200

Qy         101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
               ||||  |   | ||   || ||||| | |  ||||| || |     || ||| |
Db         201 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 253
```

RESULT 14
US-09-252-991A-10818/c
; Sequence 10818, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10818
; LENGTH: 1077
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10818

Query Match 6.5%; Score 29.8; DB 4; Length 1077;
Best Local Similarity 54.0%; Pred. No. 3.1;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy          41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
               | | || | |  |  || | |  | ||| |  |    ||||  ||||||  | |     |
Db         530 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 471

Qy         101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
               ||||  |   | ||   || ||||| | |  ||||| || |     || ||| |
Db         470 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 418
```

RESULT 15
US-09-252-991A-10865
; Sequence 10865, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10865
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10865

Query Match 6.5%; Score 29.8; DB 4; Length 1521;
Best Local Similarity 54.0%; Pred. No. 3.7;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy          41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
               | | || | |  |  || | |  | ||| |  |    ||||  ||||||  | |     |
Db         170 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 229

Qy         101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
               ||||  |   | ||   || ||||| | |  ||||| || |     || ||| |
Db         230 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 282
```

Search completed: March 18, 2004, 20:19:26
Job time : 71.4971 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	34.2	7.5	616	12	US-10-424-599-88945	Sequence 88945, A
c	2	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c	3	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c	4	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c	5	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c	6	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
	7	34	7.5	2886	9	US-09-801-368-131	Sequence 131, App
	8	33.4	7.3	7766	12	US-10-222-566-3	Sequence 3, Appli
	9	33.4	7.3	7766	14	US-10-222-162-3	Sequence 3, Appli
	10	33.4	7.3	7766	14	US-10-143-024-3	Sequence 3, Appli
c	11	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
	12	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	13	32.8	7.2	1059	9	US-09-738-626-482	Sequence 482, App
c	14	32.8	7.2	1113	9	US-09-738-626-481	Sequence 481, App
c	15	32.8	7.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
	16	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
	17	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c	18	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
c	19	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
	20	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
	21	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
	22	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c	23	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
	24	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
	25	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
c	26	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
c	27	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
c	28	31.2	6.8	2394	12	US-10-282-122A-26837	Sequence 26837, A
c	29	31	6.8	525	10	US-09-918-995-32144	Sequence 32144, A
	30	31	6.8	2751	15	US-10-398-221-3563	Sequence 3563, Ap
	31	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
c	32	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
	33	30.8	6.8	2050	14	US-10-086-510-1	Sequence 1, Appli
	34	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
	35	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appli
	36	30.6	6.7	2301	9	US-09-815-242-4844	Sequence 4844, Ap
	37	30.6	6.7	2301	9	US-09-815-242-9042	Sequence 9042, Ap
c	38	30.6	6.7	75216	14	US-10-080-170-646	Sequence 646, App
c	39	30.4	6.7	392	12	US-10-282-122A-25231	Sequence 25231, A
c	40	30.4	6.7	992	12	US-10-425-114-31747	Sequence 31747, A
c	41	30.4	6.7	25220	11	US-09-984-429-364	Sequence 364, App
	42	30.2	6.6	379	12	US-10-424-599-102524	Sequence 102524,
	43	30.2	6.6	654	12	US-10-425-114-28546	Sequence 28546, A
	44	30.2	6.6	3203	12	US-10-282-122A-24039	Sequence 24039, A
	45	30.2	6.6	3222	12	US-10-282-122A-6780	Sequence 6780, Ap

ALIGNMENTS

RESULT 1
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(616)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

Query Match 7.5%; Score 34.2; DB 12; Length 616;
Best Local Similarity 50.9%; Pred. No. 0.46;
Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
              | | |  |||  ||  || ||| || || || |||  | |    | |  | | | | |  | |
Db        556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497

Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              | | ||   || | | |    || || | |   | | |||   ||  | | |
Db        496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437
```

```
Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              || ||||||  |   |   ||| |||  ||||| ||| ||
Db        436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTA 398
```

RESULT 2
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.82;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |     |   ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 3
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.82;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |     |   ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 4
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.82;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |     |   ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 5
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora

US-10-010-390-6

Query Match 7.5%; Score 34; DB 14; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.82;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 6
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (EBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-441-736-5

Query Match 7.5%; Score 34; DB 15; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.82;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 7
US-09-801-368-131
; Sequence 131, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 131
; LENGTH: 2886
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-131

Query Match 7.5%; Score 34; DB 9; Length 2886;
Best Local Similarity 50.6%; Pred. No. 1.2;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              | ||  | | | | || | | | |||||    | |||    || |   |    |||| | |
Db       1501 AACAACAACAATAACGATAACGATAACGATAACAATAATAGCAATAATAATAATAACAAT 1560

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || | |    | | | |||  |  ||  ||||||  ||  | | || |    |||  |
Db       1561 AATAATAATAACAATAATAATAATAACAATAATAATAACAATAACAACAACAACAACAAT 1620

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
              || ||     ||| || |        || |   || | |||||
Db       1621 AATAATAATAATAGTAATAATATTAAAAATAACGTTGATAAC 1662
```

RESULT 8
US-10-222-566-3
; Sequence 3, Application US/10222566
; Publication No. US20040044192A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234USD3
; CURRENT APPLICATION NUMBER: US/10/222,566
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/125,619
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-222-566-3

Query Match 7.3%; Score 33.4; DB 12; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||     ||  ||| |  | | |  |  | |
```

```
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276

Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336

Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367

Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642
```

```
RESULT 9
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD4
;  CURRENT APPLICATION NUMBER: US/10/222,162
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-162-3

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216

Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276

Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336

Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367

Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642
```

```
RESULT 10
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD1
;  CURRENT APPLICATION NUMBER: US/10/143,024
;  CURRENT FILING DATE:  2002-05-10
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-143-024-3

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216

Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276

Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336

Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367

Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642
```

```
RESULT 11
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
;   LENGTH: 1662
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1662)
US-10-156-761-3483

  Query Match             7.3%;  Score 33.2;  DB 14;  Length 1662;
```

Best Local Similarity 52.1%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               | || | ||  | ||||| | || |  | | || || || | ||| | ||||||   |
Db         590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531

Qy         110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
                 |  ||| ||  | | |   ||   ||  | |      |  || |||   | | |||
Db         530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

Qy         170 TTGCTCTGCAAAGCGATGCCCG 191
               |||  || |   |||  | | |
Db         470 TTGAACTTCTTGGCGGCGACGG 449
```

RESULT 12
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 7.3%; Score 33.2; DB 14; Length 9025608;
Best Local Similarity 52.1%; Pred. No. 1.3e+02;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               | || | ||  | ||||| | || |  | | || || || | ||| | ||||||   |
Db     4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

Qy         110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
                 |  ||| ||  | | |   ||   ||  | |      |  || |||   | | |||
Db     4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

Qy         170 TTGCTCTGCAAAGCGATGCCCG 191
               |||  || |   |||  | | |
Db     4331901 TTGAACTTCTTGGCGGCGACGG 4331922
```

RESULT 13
US-09-738-626-482
; Sequence 482, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 482
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-482

Query Match 7.2%; Score 32.8; DB 9; Length 1059;
Best Local Similarity 47.2%; Pred. No. 1.9;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy         184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
               || | |  |   | | |  |||||  ||   |||  ||  | |||||   |||||| |
Db          76 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 135

Qy         244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
               | |   |     |  | |  ||  |  |      || ||  ||     |       |||
Db         136 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 195

Qy         304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
               | || | |   ||  | |     ||| | ||||||   || |    ||| ||| | |
Db         196 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 255

Qy         364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
               |  | ||  |||  |  ||  |  | | ||||
Db         256 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 287
```

RESULT 14
US-09-738-626-481/c
; Sequence 481, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 481

```
;   LENGTH: 1113
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-481

  Query Match             7.2%;  Score 32.8;  DB 9;  Length 1113;
  Best Local Similarity   47.2%;  Pred. No. 1.9;
  Matches  100;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | |  |    | | |  |||||  ||    |||  ||  | |||||   |||||| |
Db        972 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 913

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |      |  | |  ||  | |      || ||  ||      |       |||
Db        912 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 853

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              | || | |   ||  |  |     ||| | ||||||   || |     ||| ||| |  |
Db        852 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 793

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              |  | ||  |||  |  ||  |  | | ||||
Db        792 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 761


RESULT 15
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
;   LENGTH: 3309400
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

  Query Match             7.2%;  Score 32.8;  DB 9;  Length 3309400;
  Best Local Similarity   47.2%;  Pred. No. 1.2e+02;
  Matches  100;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | |  |    | | |  |||||  ||    |||  ||  | |||||   |||||| |
Db     451820 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 451761

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |      |  | |  ||  | |      || ||  ||      |       |||
Db     451760 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 451701

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              | || | |   ||  |  |     ||| | ||||||   || |     ||| ||| |  |
Db     451700 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 451641

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              |  | ||  |||  |  ||  |  | | ||||
Db     451640 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 451609


Search completed: March 19, 2004, 03:48:25
Job time : 272.734 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	6	US-08-233-642A-56	Sequence 56, Appl
2	456	100.0	456	23	US-09-543-407-1	Sequence 1, Appli
3	412.8	90.5	456	23	US-09-543-407-19	Sequence 19, Appl
4	406.4	89.1	456	23	US-09-543-407-17	Sequence 17, Appl
5	404.8	88.8	456	23	US-09-543-407-11	Sequence 11, Appl
6	404.8	88.8	456	23	US-09-543-407-13	Sequence 13, Appl
7	404.8	88.8	456	23	US-09-543-407-23	Sequence 23, Appl
8	404.8	88.8	456	23	US-09-543-407-25	Sequence 25, Appl
9	403.2	88.4	456	23	US-09-543-407-15	Sequence 15, Appl
10	401.6	88.1	456	23	US-09-543-407-21	Sequence 21, Appl
11	401.6	88.1	456	23	US-09-543-407-27	Sequence 27, Appl
12	398.4	87.4	456	23	US-09-543-407-29	Sequence 29, Appl
13	331.8	72.8	361	6	US-08-233-642A-54	Sequence 54, Appl
14	254.4	55.8	456	23	US-09-543-407-3	Sequence 3, Appli
15	254.4	55.8	477	17	US-09-252-691-172	Sequence 172, App
16	254.4	55.8	477	17	US-09-252-691C-172	Sequence 172, App
17	254.4	55.8	477	51	US-10-417-886-172	Sequence 172, App
18	254.4	55.8	675	13	US-08-978-878-3	Sequence 3, Appli
19	254.4	55.8	675	31	US-09-741-873B-3	Sequence 3, Appli
20	252	55.3	453	92	US-60-352-946-1	Sequence 1, Appli
21	252	55.3	453	101	US-60-444-371-1	Sequence 1, Appli
22	210.4	46.1	396	13	US-08-978-878-1	Sequence 1, Appli
23	185.6	40.7	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
C 28	44	9.6	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
29	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
30	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
31	37.4	8.2	1304	51	US-10-437-963-19773	Sequence 19773, A
32	35.8	7.9	561	17	US-09-252-691-171	Sequence 171, App
33	35.8	7.9	561	17	US-09-252-691C-171	Sequence 171, App
34	35.8	7.9	561	51	US-10-417-886-171	Sequence 171, App
C 35	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 24330, A
C 36	35.4	7.8	3412	1	PCT-US01-08631-24948	Sequence 24948, A
37	34.6	7.6	438	31	US-09-739-449-3383	Sequence 3383, Ap
38	34.6	7.6	438	33	US-09-803-110-3383	Sequence 3383, Ap
39	34.6	7.6	456	23	US-09-543-407-4	Sequence 4, Appli
C 40	34.6	7.6	29058	21	US-09-514-000-342	Sequence 342, App
41	34.6	7.6	187892	31	US-09-739-449-208	Sequence 208, App
42	34.6	7.6	187892	33	US-09-803-110-208	Sequence 208, App
C 43	34.2	7.5	385	18	US-09-333-535A-288	Sequence 288, App
C 44	34.2	7.5	385	25	US-09-615-606A-909	Sequence 909, App
C 45	34.2	7.5	396	64	US-60-089-524-4522	Sequence 4522, Ap

ALIGNMENTS

RESULT 1
US-08-233-642A-56
; Sequence 56, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-642A-56

Query Match 100.0%; Score 456; DB 6; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.9e-135;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.

```
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-135;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             90.5%;  Score 412.8;  DB 23;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 2.2e-121;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| ||| |  ||||||  | |  | |  || ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             89.1%;  Score 406.4;  DB 23;  Length 456;
  Best Local Similarity   93.2%;  Pred. No. 2.5e-119;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db          61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy         121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
               |||||||       |   ||   |     | | ||  | |    || | ||| ||||||
Db         121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy         181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy         241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy         301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy         361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy         421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
               ||||||||||||||||||||||||||||||||||
Db         421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy           1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy          61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy         121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy         181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
```

```
Qy         241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy         301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy         361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               |||||||||||||||||||| |           | | || | ||  |    | | |  |
Db         361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy         421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
               |    ||  ||||||||||||||||||||||||||||
Db         421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy           1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy          61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy         121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy         181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy         241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy         301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db         301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy         361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db         361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 8.3e-119;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 8.3e-119;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             88.4%;  Score 403.2;  DB 23;  Length 456;
  Best Local Similarity   92.8%;  Pred. No. 2.7e-118;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | |    | |      |   ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             88.1%;  Score 401.6;  DB 23;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 8.8e-118;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |  |      | || |   || | |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             88.1%;  Score 401.6;  DB 23;  Length 456;
  Best Local Similarity   92.5%;  Pred. No. 8.8e-118;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                |  |||  |    |  |  ||  |   |     |||  ||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
                |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

  Query Match             87.4%;  Score 398.4;  DB 23;  Length 456;
  Best Local Similarity   92.1%;  Pred. No. 9.4e-117;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
                ||||||  || |||  |   |                |     | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
                |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-642A-54

  Query Match             72.8%;  Score 331.8;  DB 6;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 2.2e-95;
  Matches  333;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
                |||||  |||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-3

  Query Match           55.8%;  Score 254.4;  DB 23;  Length 456;
  Best Local Similarity  72.4%;  Pred. No. 1.8e-70;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  ||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match           55.8%;  Score 254.4;  DB 17;  Length 477;
  Best Local Similarity  72.4%;  Pred. No. 1.8e-70;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| |||||||||    | ||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||     |  |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||| |||||||||| ||||||||   || |||| | |||||||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  ||||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||    ||| |||| | ||| |  ||    || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| |||| |||  | ||||| || || |  || |  || | ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  |||||||  | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


Search completed: March 19, 2004, 03:01:34
Job time : 2652.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	254.4	55.8	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	210.4	46.1	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33.4	7.3	3990	5	US-09-830-230A-629	Sequence 629, App
	4	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
	5	31.4	6.9	1588	6	US-10-767-701-14522	Sequence 14522, A
c	6	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
c	7	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
	8	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
	9	29.4	6.4	1355	6	US-10-045-674A-593	Sequence 593, App
c	10	29.4	6.4	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	11	29.4	6.4	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	12	29.4	6.4	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	13	29.4	6.4	1400	7	US-60-545-213-6407	Sequence 6407, Ap
c	14	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	15	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
c	16	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	17	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	18	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	19	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
	20	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	21	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	22	28.2	6.2	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	23	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
c	24	28.2	6.2	3416	6	US-10-100-683-3397	Sequence 3397, Ap
c	25	28.2	6.2	8206	6	US-10-021-698A-3529	Sequence 3529, Ap
	26	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
c	27	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
c	28	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
	29	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
	30	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
c	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	32	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
	33	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
	34	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	35	27.6	6.1	3035	6	US-10-786-892-279	Sequence 279, App
	36	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	37	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	38	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	39	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	40	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
	41	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
	42	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
	43	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
	44	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli
c	45	27.2	6.0	2880	6	US-10-100-683-5057	Sequence 5057, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 55.8%; Score 254.4; DB 5; Length 675;
Best Local Similarity 72.4%; Pred. No. 2.6e-77;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||  |||||    |||  |||||||||    ||     |||||  ||  ||  |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||  ||  |||||   |  |||  |  ||   |  ||  ||  |||||  ||    |||  |||  |  ||  |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||  ||   |  ||  ||||||||||   |||||  ||  ||  ||  ||  ||     ||  |  |||  |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||||   ||     |||  |  |||||  ||||||  |  |||   |||  |   |||   |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||  |||||  ||||||||  ||  |||  |  |||||  |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873C
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873C-1

  Query Match           46.1%;  Score 210.4;  DB 5;  Length 396;
  Best Local Similarity  70.7%;  Pred. No. 2.3e-62;
  Matches  280;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||  ||  |||||  ||  |  |||||||||||||  ||    ||  |||||  ||||  |  ||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  |  ||   |  |||  ||||||  |||||||||   ||  ||||  ||||  ||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||  |||||    |||  |||||||||    ||     |||||  ||  ||  |||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||  ||  |||||   |  |||  |  ||   |  ||  ||  |||||  ||    |||  |||  |  ||  |
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||  ||   |  ||  ||||||||||   |||||  ||  ||  ||  ||     ||  |  |||  |||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||||   ||     |||  |  |||||  ||||||  |  |||   |||  |   |||   |||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||  |||||  ||||||||  ||  |||  |  |||||  |||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396


RESULT 3
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Lyme Disease Vaccines
;  FILE REFERENCE: PB481US
;  CURRENT APPLICATION NUMBER: US/09/830,230A
;  CURRENT FILING DATE:  2001-09-27
;  PRIOR APPLICATION NUMBER: PCT/US98/12718
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/057,483
;  PRIOR FILING DATE: 1997-09-03
;  PRIOR APPLICATION NUMBER: 60/053,344
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/053,377
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/050,359
;  PRIOR FILING DATE: 1997-06-20
;  NUMBER OF SEQ ID NOS: 756
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
;   LENGTH: 3990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1139)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1143)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1148)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1210)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1244)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1247)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1250)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1251)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629
```

Query Match 7.3%; Score 33.4; DB 5; Length 3990;
Best Local Similarity 47.4%; Pred. No. 0.25;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db 3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy 217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db 3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy 277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db 3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 3612

Qy 337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
Db 3613 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 3643

RESULT 4
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
; APPLICANT: Neubauer, Antonie
; APPLICANT: Ziegler, Christina
; TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
; FILE REFERENCE: 1/1372
; CURRENT APPLICATION NUMBER: US/10/624,149A
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 60/403,282
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: DE 10233064
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: DE 10317008
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
; LENGTH: 150223
; TYPE: DNA
; ORGANISM: Equine herpesvirus 1
; PUBLICATION INFORMATION:
; AUTHORS: Telford,E.A.
; AUTHORS: Watson, M.S.
; AUTHORS: McBride, K.
; AUTHORS: Davison, A.J.
; TITLE: The DNA sequence of equine herpesvirus-1
; JOURNAL: Virology
; VOLUME: 189
; ISSUE: 1
; PAGES: 304-316
; DATE: JUL-1992
; DATABASE ACCESSION NUMBER: NC 001491, NCBI
; DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 2.9;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
Db 95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy 233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
Db 95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy 293 GAAATAATGCCACCATCGACCAGTGG 318
Db 95662 GAGAGCTTATTCACAACGAGCTGTTG 95687

RESULT 5
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
; LENGTH: 1588
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

Query Match 6.9%; Score 31.4; DB 6; Length 1588;
Best Local Similarity 44.5%; Pred. No. 0.79;
Matches 125; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Qy 40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
Db 455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy 100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
Db 515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy 160 AACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
Db 575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy 220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
Db 635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy 280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAA 320
Db 695 GCTGGTAGCTTTGGCGGTGGTTCCACTGGCGCCGAGAAGGA 735

RESULT 6
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.3;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
              |||   | | | | || ||   | | || |  || | |||    || |||| ||||||||
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
              | | |   | |  ||| |  |   |  |||  |  ||||||||
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583
```

RESULT 7
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
; LENGTH: 24900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(24900)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

Query Match 6.5%; Score 29.6; DB 6; Length 24900;
Best Local Similarity 52.4%; Pred. No. 12;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||  | || ||  | || | ||  |      |   | || || |  | | |||||
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |  |||||  |  | | |||||   | ||||| ||| | | | |     || |    |||
Db       9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929

Qy        327 AAAC 330
              |||
Db       9928 AATC 9925
```

RESULT 8
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

Query Match 6.4%; Score 29.4; DB 6; Length 300;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||    |     ||||  | |     | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
                ||||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233
```

RESULT 9
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
US-10-045-674A-593

Query Match 6.4%; Score 29.4; DB 6; Length 1355;
Best Local Similarity 25.2%; Pred. No. 3.6;
Matches 39; Conservative 50; Mismatches 66; Indels 0; Gaps 0;

```
Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137
              :||:||:||: ::  :  :  : ::||:||:  ::::   ||:   | :::   :   |
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197
              :  :: :  :  :||::::| :: :| :|   |:| :    |:|     ||   :
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              : |:  : |:|| ||: |:| :| : |:|  ||:|
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYG 922
```

RESULT 10
US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18

```
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2134
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2134

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 3.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   | | || || |   |||| || ||   || |   ||||
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
                |  ||   ||     |     | || ||||    | | |     |         ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   |||  |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||    ||  |   |  ||  | |     |  | ||| |
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 11
US-60-545-213-2135/c
; Sequence 2135, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2135
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2135

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 3.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   | | || || |   |||| || ||   || |   ||||
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
                |  ||   ||     |     | || ||||    | | |     |         ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   |||  |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||    ||  |   |  ||  | |     |  | ||| |
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 12
US-60-545-213-6406/c
; Sequence 6406, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6406
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6406

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 3.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   | | || || |   |||| || ||   || |   ||||
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
                |  ||   ||     |     | || ||||    | | |     |         ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   |||  |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||    ||  |   |  ||  | |     |  | ||| |
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 13
US-60-545-213-6407/c
; Sequence 6407, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6407
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6407

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 3.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   | | || || |   |||| || ||   || |   ||||
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
                |  ||   ||     |     | || ||||    | | |     |         ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   |||  |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||    ||  |   |  ||  | |     |  | ||| |
```

```
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 14
US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-117

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 249;
  Best Local Similarity   52.5%;  Pred. No. 1.9;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||     || || ||| |    | |  | |   || | | |||
Db        180 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 121

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||     |  |  ||  || || || ||| |  | |||    |||| | | |    || |
Db        120 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 61

Qy        415 CA 416
              |
Db         60 GA 59


RESULT 15
US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 757
;   TYPE: DNA
;   ORGANISM: E. Coli
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(757)
;   OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 757;
  Best Local Similarity   52.5%;  Pred. No. 3.2;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||     || || ||| |    | |  | |   || | | |||
Db        264 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 323

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||     |  |  ||  || || || ||| |  | |||    |||| | | |    || |
Db        324 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 383

Qy        415 CA 416
              |
Db        384 GA 385


Search completed: March 19, 2004, 03:09:55
Job time : 57.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	255	55.9	680	12	BJ618688	BJ618688 BJ618688
c	2	37.4	8.2	484	9	AV430994	AV430994 AV430994
	3	35.2	7.7	813	28	BZ204853	BZ204853 CH230-390
	4	34.6	7.6	702	12	BG441624	BG441624 GA__Ea001
	5	34.6	7.6	937	12	BG444255	BG444255 GA__Ea002
	6	34.2	7.5	296	9	AV058630	AV058630 AV058630
	7	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
	8	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	9	34	7.5	234	10	BB581987	BB581987 BB581987
c	10	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	11	34	7.5	811	28	BZ469000	BZ469000 BOOAO58TF
c	12	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	13	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	14	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
	15	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	16	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	17	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	18	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
c	19	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	20	33.4	7.3	505	6	AL809250	Al809250 Triticum
	21	33.4	7.3	679	29	CC961526	CC961526 BOIGS10TF
c	22	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
c	23	33.4	7.3	699	12	BI959328	BI959328 HVSMEn001
c	24	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF
c	25	33.4	7.3	709	28	BH690914	BH690914 BOHXR64TF
	26	33.4	7.3	715	13	BU444683	BU444683 603765715
c	27	33.4	7.3	785	28	BH578793	BH578793 BOGLY20TR
c	28	33.4	7.3	815	28	BH461716	BH461716 BOHIV03TR
c	29	33.4	7.3	827	28	BH438448	BH438448 BOGVQ28TR
c	30	33.4	7.3	830	28	BH715216	BH715216 BOMIM42TR
c	31	33.4	7.3	835	28	BH575194	BH575194 BOHJO27TF
c	32	33.4	7.3	853	28	BH578402	BH578402 BOHNV61TF
c	33	33.4	7.3	1026	28	BZ458784	BZ458784 BONHB14TF
	34	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
	35	33.2	7.3	625	10	BF501499	BF501499 AT16857.5
	36	33.2	7.3	629	9	AU295325	AU295325 AU295325
	37	33.2	7.3	788	9	AU139321	AU139321 AU139321
c	38	33.2	7.3	1646	10	BF579120	BF579120 602096054
	39	33	7.2	456	12	BJ331339	BJ331339 BJ331339
	40	33	7.2	524	12	BJ367372	BJ367372 BJ367372
	41	33	7.2	547	12	BJ333958	BJ333958 BJ333958
	42	33	7.2	553	12	BJ336903	BJ336903 BJ336903
	43	33	7.2	569	12	BJ334047	BJ334047 BJ334047
	44	33	7.2	574	12	BJ330328	BJ330328 BJ330328
	45	33	7.2	607	12	BJ365766	BJ365766 BJ365766

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

```
          http://xenopus.nibb.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match             55.9%;  Score 255;  DB 12;  Length 680;
  Best Local Similarity   72.4%;  Pred. No. 3.8e-62;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| | ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||   | ||    | ||| |||||| |||||||||  || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138


RESULT 2
AV430994/c
LOCUS       AV430994                484 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PL027e10_r 5', mRNA sequence.
ACCESSION   AV430994
VERSION     AV430994.1  GI:8586219
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
   PUBMED   10907854
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Porphyra yezoensis"
                     /mol_type="mRNA"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PL027e10_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             8.2%;  Score 37.4;  DB 9;  Length 484;
  Best Local Similarity   57.1%;  Pred. No. 9.5;
  Matches   68;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |   | |  | | ||  ||| |||||   ||| ||||| ||||||  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
              | ||  ||   ||  |||| ||| ||||||      |  ||    |||  |  || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323


RESULT 3
BZ204853
LOCUS       BZ204853                813 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-390D21, genomic survey sequence.
ACCESSION   BZ204853
VERSION     BZ204853.1  GI:23862905
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-390D21.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 390  row: D  column: 21
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-390D21"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
```

```
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             7.7%;  Score 35.2;  DB 28;  Length 813;
  Best Local Similarity   49.0%;  Pred. No. 51;
  Matches   94;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              ||| |   | | | || |   |   ||    || | ||  |  |   ||  |  ||
Db        548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              || |   || ||       || |  | |  |||||  |     |||||| ||  |   |
Db        608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              ||   ||  |||||||| ||| |    | ||||| | |   | |||| ||| ||     |
Db        668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAAGGACAC 727

Qy        358 GGTAATAACGCC 369
              ||  | | | |
Db        728 TGTGGTTAGGAC 739


RESULT 4
BG441624
LOCUS       BG441624                 702 bp    mRNA    linear   EST 15-MAR-2001
DEFINITION  GA__Ea0014A14f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0014A14f, mRNA sequence.
ACCESSION   BG441624
VERSION     BG441624.1  GI:13351276
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 696.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0014A14f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 702;
  Best Local Similarity   46.5%;  Pred. No. 71;
  Matches  112;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | ||| | | ||  |  |   ||   ||  |   |  |   ||  | | | || ||
Db        185 AACAATGACGGGAATGCGAATGGGAATGTTGGCAGCACAGACAATCAGAGCAACACCAAT 244

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |     |||  || |   |  || | ||  |    | || | ||  ||    |||
Db        245 GACAATGTAAATGCAAATGAAGATGTTACTCAAAAGGAAAACAACGATTCAAGTTCAAAT 304

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              | |   |  ||| | ||| | || |||   || |  | |  |   || | |  |||
Db        305 GAAAATGCTGCACAGAGCAACACCAACAACAATGAAAACAATAACAATGAAAATGCGGGT 364

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              | |||    |   | | |||||      ||||  ||  | ||  ||| ||||| |
Db        365 CAGGACGGGAAGAATGCCACTCAGAGTAACACTGACAACAATGAAAATGTGGGTCAGAAT 424

Qy        421 G 421
              |
Db        425 G 425


RESULT 5
BG444255
LOCUS       BG444255                 937 bp    mRNA    linear   EST 15-MAR-2001
DEFINITION  GA__Ea0023M14f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0023M14f, mRNA sequence.
ACCESSION   BG444255
VERSION     BG444255.1  GI:13353907
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 937)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 804.
FEATURES             Location/Qualifiers
     source          1. .937
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0023M14f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 937;
  Best Local Similarity   46.5%;  Pred. No. 81;
  Matches  112;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | ||| | | ||  |  |   ||   ||  |   |  |   ||  | | | || ||
Db        185 AACAATGACGGGAATGCGAATGGGAATGTTGGCAGCACAGACAATCAGAGCAACACCAAT 244

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
```

```
Db        245 GACAATGTAAATGCAAATGAAGATGTTACTCAAAAGGAAAACAACGATTCAAGTTCAAAT 304

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        305 GAAAATGCTGCACAGAGCAACACCAACAACAATGAAAACAATAACAATGAAAATGCGGGT 364

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        365 CAGGACGGGAAGAATGCCACTCAGAGTAACACTGACAACAATGAAAATGTGGGTCAGAAT 424

Qy        421 G 421

Db        425 G 425
```

RESULT 6
AV058630
LOCUS AV058630 296 bp mRNA linear EST 23-JUN-1999
DEFINITION AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA clone 1810054H02, mRNA sequence.
ACCESSION AV058630
VERSION AV058630.1 GI:5158377
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 296)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"
/clone_lib="Mus musculus pancreas C57BL/6J adult"
ORIGIN

Query Match 7.5%; Score 34.2; DB 9; Length 296;
Best Local Similarity 53.3%; Pred. No. 63;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCG 310

Db         13 GTGGAGACCACAGTTTTTTCTAGGTCACCAAGGCTGGGTAAAGAAAAGAAGCATCAACTG 72

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370

Db         73 TCCAGTGAAGCACAAAAAATGAAGAAGAAAATGATTACCCAAAAGTTGGAATTAAGGTAG 132

Qy        371 CGCTGGTTAATCAGA 385

Db        133 AGATGCCAAAATAGA 147
```

RESULT 7
BJ334624
LOCUS BJ334624 523 bp mRNA linear EST 05-MAR-2002
DEFINITION BJ334624 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda47o06 5', mRNA sequence.
ACCESSION BJ334624
VERSION BJ334624.1 GI:19164754
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 523)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .523
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47o06"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 523;
Best Local Similarity 57.1%; Pred. No. 82;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297

Db        398 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGANGACAATAAT 457

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACT 342

Db        458 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAAT 502
```

RESULT 8
BG671547
LOCUS BG671547 1036 bp mRNA linear EST 30-APR-2001
DEFINITION DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5', mRNA sequence.
ACCESSION BG671547
VERSION BG671547.1 GI:13893646
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1036)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C.,

```
            Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D.,
            Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
   PUBMED   12060780
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1036
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNBUB05"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /clone_lib="Rat DRG Library"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 1036;
  Best Local Similarity   49.2%;  Pred. No. 1.1e+02;
  Matches   90;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy          11 TAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTC 70
               | ||||   ||  | |  |||    |||  || | | || |||| |     |||||  |
Db         168 TGAAAGATACATGAGTGACAGTGTCCGTGATTGCAGACATTGCTGTATGATGCGTCCTGT 227

Qy          71 CACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAA 130
               |   | | | |  ||  |||||  ||  ||   | |   || | |   |   ||     |
Db         228 CTGTAGGAGTCCTCGCTGGTAACGATGCCGAGAGTGCACATCGATATCGGTGGGCGGTTA 287

Qy         131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
               | |       |  |  ||   | |    | ||| |||   ||| ||||  ||| || ||
Db         288 CATCTGAAGTGGAAGAGGTGGCTCTTTCGAACTCTGACTTTGATCTGTCAAGGGACGCTT 347

Qy         191 GTA 193
               |||
Db         348 GTA 350


RESULT 9
BB581987/c
LOCUS       BB581987                 234 bp    mRNA    linear   EST 30-NOV-2000
DEFINITION  BB581987 RIKEN full-length enriched, adult male colon Mus musculus
            cDNA clone 9030005O05 5', mRNA sequence.
ACCESSION   BB581987
VERSION     BB581987.1  GI:11478531
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
            Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
            Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
            Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
            Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
            Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9030005O05"
                     /sex="male"
                     /tissue_type="colon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male colon"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 10;  Length 234;
  Best Local Similarity   51.3%;  Pred. No. 65;
  Matches   79;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          35 TCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATC 94
               | ||  ||  | |    | |   || ||  || |    ||   || | || || ||   |
Db         200 TAGTTTTTGGTAGGTCGGTTTATGCGGGTTTCTTCTTCCACCAGGCCCGCCGCCGTTTCC 141
```

```
Qy         95 ATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTT 154
              |||| |||  ||  ||   ||        |   |  || | ||||||| ||  ||  ||
Db        140 CGAACGCCGGGTGCTTTATAGCCTTGTTCAAAATATGGTAGCGCATTTACCACCACCTTT 81

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||| | ||   || | ||| |  ||  |||  ||
Db         80 CCGGTTACCAAGCCCGTGCCCGCCATTGCGCGGC 47


RESULT 10
CA711477/c
LOCUS       CA711477                500 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wdk2c.pk014.o3 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.o3 5'
            end, mRNA sequence.
ACCESSION   CA711477
VERSION     CA711477.1  GI:25433270
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone="wdk2c.pk014.o3"
                     /tissue_type="kernel"
                     /clone_lib="wdk2c"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
                     days after anthesis."
ORIGIN

  Query Match             7.5%;  Score 34;  DB 14;  Length 500;
  Best Local Similarity   54.5%;  Pred. No. 91;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
               ||  |  ||| |||||||   ||   |  | ||| ||       | | ||  | |||| |
Db         274 AGNTGNTGCAGTCGCAGCCNCCGGTCTCCCCGGCCGTNTCGAACCCGCCGGAGCTCCCCN 215

Qy          75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGAC 126
               ||| ||||| ||  || |  | ||||||||||     |    ||||   ||
Db         214 TGTGGGAGGCGGTGGGCATGACCATGGCGGCGGCGCCGGCGGTGGCCNCCAC 163


RESULT 11
BZ469000/c
LOCUS       BZ469000                811 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BOOAO58TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOAO58,
            genomic survey sequence.
ACCESSION   BZ469000
VERSION     BZ469000.1  GI:26764546
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOOAO58TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOOAO58"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 28;  Length 811;
  Best Local Similarity   48.9%;  Pred. No. 1.1e+02;
  Matches   91;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy         176 TGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCG 235
               ||  | || |     ||||| |||       |||||  |     |  | |||||  ||  ||
Db         610 TGTTAGGCAAGCATTGTAAAACTGTGTTTCCCATTTCATCTACTGTTTATGACAAATGCT 551

Qy         236 CCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
               |||  ||   ||   ||  | |   | | ||  |     || ||  | || |||| | |
Db         550 TTGATCTAATTCACTCTGATGTTTGGACTGCTCCTTGCTTGTCTAGGGATAGTTTTAAAT 491

Qy         296 ATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 355
               |   || |||  ||     |    ||  | ||| || ||    || ||  |    |||  |
Db         490 ACTTTGTCACATTCATAAATGAAAAATCGAAATACACCTGGATAACACTTATTCAAACCA 431

Qy         356 GCGGTA 361
               | ||
Db         430 AAGATA 425


RESULT 12
AA746477/c
LOCUS       AA746477                365 bp    mRNA    linear   EST 27-JAN-1998
DEFINITION  nw62e02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
            similar to TR:O00555 O00555 NEURONAL CALCIUM CHANNEL ALPHA 1A
            SUBUNIT ISOFORM A-1. ;contains element TAR1 repetitive element ;,
            mRNA sequence.
ACCESSION   AA746477
VERSION     AA746477.1  GI:2786463
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
```

```
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html
           Insert Length: 661    Std Error: 0.00
           Seq primer: -40m13 fwd. ET from Amersham
           High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1251194"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 365;
  Best Local Similarity   56.9%;  Pred. No. 91;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  ||||||   ||||||   | ||   | ||| |||  | |||  |||   |  ||||
Db        119 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGTGGCCAGGCCGGCCGGGCGCCACC 60

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |  || |  ||| ||||  ||   | ||   | | |   | | |||| |
Db         59 AGCGGCCCTCGGAGGTACCCAGGCCCGCACGGCCGACGCTACTGGCCGG 11


RESULT 13
CD374421/c
LOCUS       CD374421                  558 bp    mRNA    linear   EST 19-AUG-2003
DEFINITION  TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)
            Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION   CD374421
VERSION     CD374421.1  GI:31229939
KEYWORDS    EST.
SOURCE      Wuchereria bancrofti
  ORGANISM  Wuchereria bancrofti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Wuchereria.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
  TITLE     Toward the characterization of the Wuchereria bancrofti
            microfilarial transcriptome with comparisons to those of Brugia
            malayi and Onchocerca volvulus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Thomas B. Nutman
            Laboratory of Parasitic Diseases
            NIH
            Building 4 Room 126, Bethesda, MD 20892-0425, USA
            Email: tnutman@niaid.nih.gov
            Seq primer: pBluescript T3.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Wuchereria bancrofti"
                     /mol_type="mRNA"
                     /db_xref="taxon:6293"
                     /clone="TNWbmfC2H3"
                     /dev_stage="microfilaria"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="Wuchereria bancrofti microfilaria cDNA
                     (SAW95SjL-WbMf)"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 85,000 microfilariae
                     isolated from the blood of an infected individual from
                     Guyana, South America and converted to double-stranded
                     cDNA using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNA pol I. The library has 2.9 x 10E5
                     independent recombinants and the average insert size is
                     ~1kb. The library was constructed by Sandra J. Laney. The
                     library is available from Dr.S.A.Williams, email:
                     genome@smith.edu."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 558;
  Best Local Similarity   60.7%;  Pred. No. 1.1e+02;
  Matches   65;  Conservative    0;  Mismatches   40;  Indels    2;  Gaps    1;

Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              ||| |||| |||||||| |     || ||||| |||| |    | ||  | |
Db        161 TGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCCCAAATTCTANGCT 102

Qy        136 AGCATTTATC--AGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||| |  ||  ||||  ||  | | |||| |||||||| ||
Db        101 GACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55


RESULT 14
BU765156/c
LOCUS       BU765156                  559 bp    mRNA    linear   EST 10-OCT-2002
DEFINITION  sas23f11.y1 Gm-c1080 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1080-4869 5' similar to SW:D100_ARATH Q00874
            DNA-DAMAGE-REPAIR/TOLERATION PROTEIN DRT100 PRECURSOR ;, mRNA
            sequence.
ACCESSION   BU765156
VERSION     BU765156.1  GI:23733884
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Glycine max"
```

/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1080-4869"
/tissue_type="Roots of 8 day old 'Bragg' supernodulating mutant NTS382 seedlings"
/dev_stage="8 days old"
/lab_host="DH10B"
/clone_lib="Gm-c1080"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The mRNA was isolated from roots of 8 day old 'Bragg' supernodulating mutant NTS382 seedlings that were infected with Bradyhizobium japonicum, strain USDA 110, 72 hours prior to harvest. Dr. Gary Stacey generously donated the tissue. The roots were flash-frozen in liquid nitrogen. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An 'anchor' nucleotide (V=A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first- strand synthesis primer was then restricted by digestion with XhoI from Promega (40U/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using Sephacryl S-500 High Resolution (Pharmacia Biotech) in a 2-mm diameter column and a bed volume of approximately 1ml. The column eluent was precipitated, redissolved, and ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). This library was constucted in the laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at Northern Arizona University."

ORIGIN

Query Match 7.4%; Score 33.8; DB 13; Length 559;
Best Local Similarity 55.6%; Pred. No. 1.1e+02;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              |||||  || |  |    |||| | ||  |||||| ||| | |  || |||     ||||
Db        233 TGGGGTCGCAGGCGACGCCATACCAGCTCCGGCAACAGTACAGTGCCCGACCATGTGTTG 174

Qy        136 AGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGT 192
              |  ||     |||| |||||    |  || ||  |        | || | |||||||
Db        173 AAGATGCCCAAGTAGGGTTCGATCAGGGCAGCTTTGAAGGCCAAGAGTGCTGCCCGT 117
```

RESULT 15
CG178280
LOCUS CG178280 941 bp DNA linear GSS 21-AUG-2003
DEFINITION PUJDF79TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0641N14, genomic survey sequence.
ACCESSION CG178280
VERSION CG178280.1 GI:34069341
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 941)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUJDF79TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .941
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa0641N14"
/clone_lib="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high CoT selected genomic DNA library"
ORIGIN

Query Match 7.4%; Score 33.8; DB 29; Length 941;
Best Local Similarity 51.0%; Pred. No. 1.4e+02;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | ||||  | | || || | || || | || ||  | ||  ||    | | |||| |
Db        236 TTGCAGTCTACACATCAGTGGTCGTATATGCCACAGTTCCTGCACCGGCCTTTCCGCCGC 295

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGA 136
              || |   | | ||  |||  ||| | || |   |    | ||| |||||  ||   ||
Db        296 GGAGTACCCTCCGTCGTCACCACGACCGCCGGTCTTCGACTTGCCGGGACTGTACCCCGA 355

Qy        137 GCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGC 173
                || ||   || | ||  |     |  | || ||
Db        356 CTTCTTCTCGCCACCGCTCAACGGCGACGACTCTGGC 392
```

Search completed: March 18, 2004, 20:11:06
Job time : 2038.55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	404.8	88.8	456	6	I44909	I44909 Sequence 58
2	404.8	88.8	2067	1	SEU43280	U43280 Salmonella
3	396.8	87.0	5103	1	STAJ2301	AJ002301 Salmonell
4	396.8	87.0	22411	1	AE008749	AE008749 Salmonell
5	395.2	86.7	254050	1	AL627269	AL627269 Salmonell
c 6	395.2	86.7	301983	1	AE016840	AE016840 Salmonell
7	389.4	85.4	1048	1	STAGFBA	AJ000514 Salmonell
8	315.2	69.1	361	6	I44908	I44908 Sequence 56
9	275.2	60.4	2889	1	CSP515700	AJ515700 Citrobact
10	232.4	51.0	2920	1	CFR515701	AJ515701 Citrobact
11	229	50.2	230	1	SEU53207	U53207 Salmonella
12	217.6	47.7	4680	1	ECCSGABDG	X90754 E.coli csgG
13	217.6	47.7	10346	1	AE000205	AE000205 Escherich
14	217.6	47.7	15047	1	D90741	D90741 Escherichia
15	216	47.4	456	6	AX814811	AX814811 Sequence
16	216	47.4	648	1	ECOCSGAA	L04979 Escherichia
17	212.6	46.6	306358	1	AE016759	AE016759 Escherich
18	209.4	45.9	1711	1	AF275733	AF275733 Escherich
19	209.4	45.9	10190	1	AE005315	AE005315 Escherich
20	209.4	45.9	327773	1	AP002554	AP002554 Escherich
21	184	40.4	2883	1	ESA515702	AJ515702 Enterobac
22	155.2	34.0	10370	1	AE015131	AE015131 Shigella
23	155.2	34.0	292504	1	AE016981	AE016981 Shigella
c 24	143.2	31.4	437	1	AF237726	AF237726 Shigella
25	64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
26	54.8	12.0	19201	1	D90742	D90742 Escherichia
27	44.8	9.8	72	6	AX814798	AX814798 Sequence
c 28	39.4	8.6	2000	6	AX655393	AX655393 Sequence
c 29	38	8.3	6669	5	AF229448	AF229448 Danio rer
c 30	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
c 31	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
c 32	36	7.9	168267	2	AC146009	AC146009 Pan trogl
33	35.6	7.8	301214	1	AE016786	AE016786 Pseudomon
c 34	35	7.7	290029	1	AE017038	AE017038 Bacillus
35	34.4	7.5	547	5	AF523644	AF523644 Gavia imm
36	34.4	7.5	151828	2	AC146008	AC146008 Pan trogl
c 37	34.4	7.5	183711	2	AP001445	AP001445 Homo sapi
c 38	34.4	7.5	211981	9	AC009716	AC009716 Homo sapi
39	34.2	7.5	1458	1	AF173043	AF173043 Orientia
40	34.2	7.5	1803	3	AF325159	AF325159 Plasmodiu
41	34.2	7.5	1803	3	AF325160	AF325160 Plasmodiu
42	34.2	7.5	1803	3	AF325161	AF325161 Plasmodiu
c 43	34.2	7.5	104680	9	AC024591	AC024591 Homo sapi
c 44	34.2	7.5	189756	9	AC092327	AC092327 Homo sapi
45	34	7.5	956	8	AK058943	AK058943 Oryza sat

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 7e-104;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match             88.8%;  Score 404.8;  DB 1;  Length 2067;
  Best Local Similarity   93.0%;  Pred. No. 7.1e-104;
```

```
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||| |          | | || | ||  |    | | |  |
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match             87.0%;  Score 396.8;  DB 1;  Length 5103;
  Best Local Similarity   91.9%;  Pred. No. 1.3e-101;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |   | | | |
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA

COMMENT COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES Location/Qualifiers

source 1. .22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

gene 434. .1308
/gene="phoH"
/note="synonym: STM1126"

RBS 434. .439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"

CDS 454. .1308
/gene="phoH"
/note="similar to E. coli PhoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=11
/product="PhoB-dependent ATP-binding pho regulon component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
QRTLHAYS"

gene complement(1414. .2303)
/gene="STM1127"

CDS complement(1414. .2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29% identity in aa 15 - 264"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
QARESKLRTMNALDMTK"

RBS complement(2298. .2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"

gene complement(2580. .4088)
/gene="STM1128"

CDS complement(2580. .4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22% identity in aa 7 - 478"
/codon_start=1
/transl_table=11
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPPFRK
LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"

RBS complement(4083. .4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"

gene complement(4413. .5099)
/gene="STM1129"

CDS complement(4413. .5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 - 226"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
WFSHAVKR"

RBS complement(5094. .5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"

gene 5599. .6759
/gene="STM1130"

CDS 5599. .6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41% identity in aa 40 - 403"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVPLNDELYVFG
GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"

gene 6793. .7497
/gene="STM1131"

RBS 6793. .6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

 Query Match             87.0%;  Score 396.8;  DB 1;  Length 22411;
 Best Local Similarity   91.9%;  Pred. No. 1.3e-101;
 Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269              254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEPLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             86.7%;  Score 395.2;  DB 1;  Length 254050;
  Best Local Similarity   91.7%;  Pred. No. 3.9e-101;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |        | | || | ||  |    | | |  |
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.7%;  Score 395.2;  DB 1;  Length 301983;
Best Local Similarity   91.7%;  Pred. No. 3.9e-101;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |         | | || | ||  |    | | |  |
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             85.4%;  Score 389.4;  DB 1;  Length 1048;
  Best Local Similarity   91.0%;  Pred. No. 1.7e-99;
  Matches  414;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTA 455
              |    ||  ||||||||  |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             69.1%;  Score 315.2;  DB 6;  Length 361;
  Best Local Similarity   99.1%;  Pred. No. 2.1e-78;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              |||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           60.4%;  Score 275.2;  DB 1;  Length 2889;
  Best Local Similarity  76.8%;  Pred. No. 4.9e-67;
  Matches  350;  Conservative    0;  Mismatches  103;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |    ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  ||||||||||||||||  ||   |  |||||| |||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||    ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || |||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||| ||||| ||  |   |  |   | | || | ||| |    ||| |  |
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMPTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
  Query Match             51.0%;  Score 232.4;  DB 1;  Length 2920;
  Best Local Similarity   71.1%;  Pred. No. 7.3e-55;
  Matches  324;  Conservative    0;  Mismatches  126;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| ||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||    ||| ||||  ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||    ||||| || ||   |||   ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||||  ||| | ||||| |||   ||||  ||||  || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| ||||| |||||||| || |||||  | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                 ||||  ||| ||||||||  |           |   ||  |||  | |  ||  |  |
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

RESULT 11
SEU53207

```
LOCUS       SEU53207                 230 bp    DNA     linear   BCT 01-MAY-1996
DEFINITION  Salmonella enteritidis SEF17 fimbrin (agfA) gene, partial cds.
ACCESSION   U53207
VERSION     U53207.1  GI:1293677
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 230)
  AUTHORS   Cox,J.M., Eglezos,S. and Woolcock,J.B.
  TITLE     Virulence of Salmonella enteritidis in chickens correlates with
            colony morphology and expression of SEF17 fimbriae
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230)
  AUTHORS   Cox,J.M., Eglezos,S. and Woolcock,J.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-1996) J.M. Cox, Food Science and Technology, The
            University of New South Wales, Sydney, NSW 2052, Australia
FEATURES             Location/Qualifiers
     source          1. .230
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="SE30"
                     /db_xref="taxon:592"
                     /note="amplification product"
     gene            1. .230
                     /gene="agfA"
     CDS             <1. .>230
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="SEF17 fimbrin"
                     /protein_id="AAA98671.1"
                     /db_xref="GI:1293678"
                     /translation="GNHXGGGNSSGPDSTLSIYQYGSANAALALQSDARKSETTITQS
                     GYGNGADVGQGADNSTIELTQNGFRNNATIDQ"
ORIGIN
```

```
  Query Match             50.2%;  Score 229;  DB 1;  Length 230;
  Best Local Similarity   99.6%;  Pred. No. 6.7e-54;
  Matches  229;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         88 GGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 147
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTAATCATAANGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 60

Qy        148 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACC 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACC 120

Qy        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 180

Qy        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 230
```

RESULT 12
ECCSGABDG

```
LOCUS       ECCSGABDG                4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
```

```
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match             47.7%;  Score 217.6;  DB 1;  Length 4680;
  Best Local Similarity   67.3%;  Pred. No. 1.2e-50;
  Matches  307;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||   ||| |  |          ||   ||          | | |
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184
```

```
RESULT 13
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFPSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
```

```
                VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene       1513. .1926
                /gene="ycdV"
                /note="synonym: b1031"
     CDS        1513. .1926
                /gene="ycdV"
                /function="putative structure; Not classified"
                /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                P05759"
                /codon_start=1
                /transl_table=11
                /product="putative ribosomal protein"
                /protein_id="AAC74116.1"
                /db_xref="GI:1787269"
                /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene       complement(2111. .2198)
                /gene="serX"
                /note="synonym: b1032"
     tRNA       complement(2111. .2198)
                /gene="serX"
                /product="tRNA-Ser"
                /note="anticodon: GGA; CG Site No. 17878"
                /function="RNA; tRNA"
     promoter   2269. .2296
                /note="factor Sigma70; predicted +1 start at 1096980"
     promoter   2338. .2365
                /note="factor Sigma70; predicted +1 start at 1097049"
     promoter   2355. .2383
                /note="factor Sigma70; predicted +1 start at 1097067"
     gene       2393. .3370
                /gene="ycdW"
                /note="synonym: b1033"
     CDS        2393. .3370
                /gene="ycdW"
                /function="putative enzyme; Not classified"
                /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                to 177 residues of an approx. 312 aa protein YPRA_CORGL
                SW: P45637"
                /codon_start=1
                /transl_table=11
                /product="putative dehydrogenase"
                /protein_id="AAC74117.1"
                /db_xref="GI:1787270"
                /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter   3278. .3308
                /gene="ycdW"
                /note="factor Sigma70; predicted +1 start at 1097992"
     promoter   3383. .3410
                /note="factor Sigma70; predicted +1 start at 1098094"
     promoter   3396. .3423

  Query Match          47.7%;  Score 217.6;  DB 1;  Length 10346;
  Best Local Similarity  67.3%;  Pred. No. 1.2e-50;
  Matches  307;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||

Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                ||||  ||    ||| |  |   |          ||   ||         | |  |
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448


RESULT 14
D90741
LOCUS       D90741                15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
```

```
            Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
            Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
            Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
            Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
            Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
           Headed by:
            Name: Takashi Horiuchi
            Address: National Institute of Basic Biology, Okazaki, 444, Japan
            E-mail: kishori@nibb.ac.jp
           Information operator:
            Name: Hirotada Mori
            Address: NARA Institute of Science and Technology,
                  Ikoma, 630-01, Japan
            E-mail: hmori@gtc.aist-nara.ac.jp
           URL:
            The Japan E. coli genome database
            http:bsw3.aist-nara.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .15047
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:83333"
                     /map="23.7-24.0 min"
                     /clone="Kohara clone #231"
                     /note="Nucleotide position 1097101-1112147 from the
                     initiation site of ThrA (0 min.).~This clone is from
                     Kohara lambda miniset library."
     CDS             <1. .276
                     /note="unnamed protein product; ORF_ID:o230#5"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35822.1"
                     /db_xref="GI:4062605"
                     /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
                     WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     CDS             1152. .2090
                     /note="ORF_ID:o231#1
                     similar to SwissProt Accession Number P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical 33.0 kd protein in proB-proA
                     intergenic region."
                     /protein_id="BAA35823.1"
                     /db_xref="GI:4062606"
                     /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
                     WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
                     AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
                     PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
                     GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
                     AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     CDS             2145. .2882
                     /note="unnamed protein product; ORF_ID:o231#2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35824.1"
                     /db_xref="GI:4062607"
                     /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                     EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                     PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                     SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
                     ILNVSPRRLLNFLESRGMAPIAEFADL"
     CDS             2906. .3460
                     /note="ORF_ID:o231#3
                     similar to PIR Accession Number F64035"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein HI1543"
                     /protein_id="BAA35825.1"
                     /db_xref="GI:4062608"
                     /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                     DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                     ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                     RTMAPLTRDAISAMWDELEEDSEE"
     CDS             3514. .4053
                     /note="unnamed protein product; ORF_ID:o231#4"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35826.1"
                     /db_xref="GI:4062609"
                     /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                     SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                     IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                     LAARSAKTAHREQEIKNKA"
     gene            complement(4117. .4950)
                     /gene="csgG"
     CDS             complement(4117. .4950)
                     /gene="csgG"
                     /note="ORF_ID:o231#5
                     similar to SwissProt Accession Number P52103"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgG precursor."
                     /protein_id="BAA35827.1"
                     /db_xref="GI:4062610"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(4977. .5393)
                     /gene="csgF"
     CDS             complement(4977. .5393)
                     /gene="csgF"
                     /note="ORF_ID:o231#6
                     similar to SwissProt Accession Number P52104"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgF precursor."
                     /protein_id="BAA35828.1"
                     /db_xref="GI:4062611"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(5418. .5807)
                     /gene="csgE"
     CDS             complement(5418. .5807)
                     /gene="csgE"
                     /note="ORF_ID:o231#7
                     similar to SwissProt Accession Number P52105"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgE precursor."
                     /protein_id="BAA35829.1"
                     /db_xref="GI:4062612"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(5812. .6462)
                     /gene="csgD"
     CDS             complement(5812. .6462)
                     /gene="csgD"
                     /note="ORF_ID:o231#8
                     similar to SwissProt Accession Number P52106"
                     /codon_start=1
                     /transl_table=11
                     /product="Probable csgAB operon transcriptional regulatory
                     protein."
                     /protein_id="BAA35830.1"
                     /db_xref="GI:4062613"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
```

```
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLPISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            7217. .7672
                     /gene="csgB"
     CDS             7217. .7672

  Query Match           47.7%;  Score 217.6;  DB 1;  Length 15047;
  Best Local Similarity  67.3%;  Pred. No. 1.2e-50;
  Matches  307;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||| ||||| || || || || ||    || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||    ||| |  |   |           ||   ||         | |  |
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 15
AX814811
LOCUS       AX814811                456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           47.4%;  Score 216;  DB 6;  Length 456;
  Best Local Similarity  67.1%;  Pred. No. 3.3e-50;
  Matches  306;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||| ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||    ||| |  |   |           ||   ||         | |  |
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


Search completed: March 18, 2004, 15:02:42
Job time : 2312.32 secs
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	3	AAC64622	Aac64622 AgfA::PT3
2	404.8	88.8	456	2	AAQ87467	Aaq87467 AgfA sequ
3	404.8	88.8	456	2	AAT74142	Aat74142 Salmonell
4	404.8	88.8	456	3	AAC64617	Aac64617 Salmonell
5	376	82.5	456	3	AAC64623	Aac64623 AgfA::PT3
6	361.6	79.3	456	3	AAC64626	Aac64626 AgfA::PT3
7	355.2	77.9	456	3	AAC64625	Aac64625 AgfA::PT3
8	353.6	77.5	456	3	AAC64628	Aac64628 AgfA::PT3
9	353.6	77.5	456	3	AAC64629	Aac64629 AgfA::PT3
10	352	77.2	456	3	AAC64624	Aac64624 AgfA::PT3
11	350.4	76.8	456	3	AAC64630	Aac64630 AgfA::PT3
12	350.4	76.8	456	3	AAC64627	Aac64627 AgfA::PT3
13	347.2	76.1	456	3	AAC64631	Aac64631 AgfA::PT3
14	315.2	69.1	361	2	AAQ73066	Aaq73066 AgfA sequ
15	315.2	69.1	361	2	AAT74141	Aat74141 Salmonell
16	216	47.4	456	3	AAC64619	Aac64619 Escherich
17	216	47.4	456	9	ACF36153	Acf36153 E. coli C
18	211.6	46.4	646	2	AAQ62647	Aaq62647 Fibronect
19	163.2	35.8	369	2	AAQ62646	Aaq62646 FNB curli
20	52.2	11.4	78	3	AAC64610	Aac64610 AgfA (SEF
c 21	49.4	10.8	78	3	AAC64609	Aac64609 AgfA (SEF
22	48	10.5	48	3	AAC64621	Aac64621 Leishmani
23	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
24	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
c 25	40.8	8.9	78	3	AAC64605	Aac64605 SefA (SEF
26	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
27	40.6	8.9	78	3	AAC64606	Aac64606 SefA (SEF
c 28	39.4	8.6	2000	7	ADA71938	Ada71938 Rice gene
29	39	8.6	100	7	ACD68808	Acd68808 E. coli K
30	36	7.9	557	6	ABQ43459	Abq43459 Oligonucl
c 31	36	7.9	557	6	ABQ43458	Abq43458 Oligonucl
32	36	7.9	646	6	ABQ46953	Abq46953 Oligonucl
c 33	36	7.9	646	6	ABQ46952	Abq46952 Oligonucl
34	35.4	7.8	100	7	ACD68807	Acd68807 E. coli K
c 35	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c 36	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
37	34.6	7.6	456	3	AAC64620	Aac64620 Escherich
38	34	7.5	456	3	AAC64618	Aac64618 Salmonell
c 39	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c 40	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c 41	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c 42	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c 43	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c 44	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c 45	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a

ALIGNMENTS

RESULT 1
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

```
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-137;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag= a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 2;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.1e-120;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |           | | || | ||  |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
AAT74142
```

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 2;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.1e-120;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.1e-120;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 82.5%; Score 376; DB 3; Length 456;
Best Local Similarity 89.0%; Pred. No. 2.4e-111;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                  ||    | ||||  ||||  |   |  | |  |    | | |  |
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match           79.3%;  Score 361.6;  DB 3;  Length 456;
  Best Local Similarity  87.1%;  Pred. No. 1.1e-106;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |       | | || | ||  |    | | |   |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 77.9%; Score 355.2; DB 3; Length 456;
Best Local Similarity 86.2%; Pred. No. 1.3e-104;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |   ||   |    | | ||  | |    || | ||| |||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||||| |          | | || | ||  |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 8
AAC64628
ID AAC64628 standard; DNA; 456 BP.
XX
AC AAC64628;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36352.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 4.4e-104;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||  |||      ||  ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||  |    ||  |  |||  |          | |  || |  ||  |    | | |  |
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 9
AAC64629
ID AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 4.4e-104;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||  |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 1.5e-103;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | |   |  ||   |   ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |           | | || | ||  |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 4.8e-103;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||  |     | | ||  |   |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |     | | || | ||  |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 4.8e-103;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |  |       | ||  |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match           76.1%;  Score 347.2;  DB 3;  Length 456;
  Best Local Similarity  85.1%;  Pred. No. 5.3e-102;
  Matches  388;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |  |                 |  |  ||  |||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||  |           | | || | | |  |  | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match            69.1%;  Score 315.2;  DB 2;  Length 361;
  Best Local Similarity  99.1%;  Pred. No. 1.1e-91;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              |||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag= b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag= c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag= d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag= e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match            69.1%;  Score 315.2;  DB 2;  Length 361;
  Best Local Similarity  99.1%;  Pred. No. 1.1e-91;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              |||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320


Search completed: March 18, 2004, 09:11:55
```

Job time : 289.405 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	404.8	88.8	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	315.2	69.1	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	33.4	7.3	7766	4	US-09-125-619-3	Sequence 3, Appli
	6	33.4	7.3	14770	4	US-09-220-132-30	Sequence 30, Appl
c	7	32.2	7.1	1122	4	US-09-489-039A-3211	Sequence 3211, Ap
c	8	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	9	31.2	6.8	2436	4	US-09-540-236-492	Sequence 492, App
c	10	31.2	6.8	65792	4	US-09-596-002-31	Sequence 31, Appl
	11	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
c	12	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	13	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	14	30	6.6	1341	4	US-09-350-756-2	Sequence 2, Appli
	15	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	16	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	17	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
c	18	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
c	19	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	20	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	21	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	22	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	23	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	24	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, App
c	25	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, App
c	26	29.4	6.4	3494	4	US-09-023-655-1006	Sequence 1006, Ap
c	27	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
	28	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
	29	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	30	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	31	29	6.4	534	4	US-09-489-039A-2195	Sequence 2195, Ap
	32	29	6.4	1200	4	US-09-134-000C-1704	Sequence 1704, Ap
	33	29	6.4	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
c	34	29	6.4	13794	4	US-08-956-171E-54	Sequence 54, Appl
c	35	28.8	6.3	654	4	US-09-252-991A-4506	Sequence 4506, Ap
	36	28.8	6.3	720	4	US-09-328-352-354	Sequence 354, App
	37	28.8	6.3	1662	1	US-08-565-386-2	Sequence 2, Appli
	38	28.8	6.3	4529	1	US-08-565-386-1	Sequence 1, Appli
c	39	28.8	6.3	8906	4	US-09-027-169-5	Sequence 5, Appli
c	40	28.6	6.3	432	4	US-09-252-991A-4473	Sequence 4473, Ap
	41	28.4	6.2	1026	3	US-07-751-891B-24	Sequence 24, Appl
	42	28.4	6.2	4188	3	US-07-751-891B-2	Sequence 2, Appli
	43	28.4	6.2	4242	3	US-07-705-490-2	Sequence 2, Appli
	44	28.4	6.2	5222	3	US-07-751-891B-23	Sequence 23, Appl
	45	28.2	6.2	357	4	US-09-489-039A-1118	Sequence 1118, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043:403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 4.5e-131;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             69.1%;  Score 315.2;  DB 1;  Length 361;
  Best Local Similarity   99.1%;  Pred. No. 6.5e-100;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              |||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320
```

```
RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
;  GENERAL INFORMATION:
;    APPLICANT:  Kim, Jihyun Francis
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;    TITLE OF INVENTION:  ERWINIA AMYLOVORA AND ITS USE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,927
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,108
;      FILING DATE:  06-AUG-1977
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1581
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
```

```
;       TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1344 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.13;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.13;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-09-125-619-3

  Query Match             7.3%;  Score 33.4;  DB 4;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 0.56;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||    ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |    |  | ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    |  ||| |          | |||  |||||  |      |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||       | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 6
US-09-220-132-30
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  TITLE OF INVENTION:  METHODS AND COMPOSITONS FOR THE IDENTIFICATION AND ASSESSMENT
;  TITLE OF INVENTION:  OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCH
;  FILE REFERENCE:  07334-074001
;  CURRENT APPLICATION NUMBER: US/09/220,132
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  US 60/079,303
;  PRIOR FILING DATE:  1998-03-25
;  PRIOR APPLICATION NUMBER:  US 60/068,821
;  PRIOR FILING DATE:  1997-12-24
;  NUMBER OF SEQ ID NOS:  191
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 30
;   LENGTH: 14770
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-220-132-30

  Query Match             7.3%;  Score 33.4;  DB 4;  Length 14770;
  Best Local Similarity   48.7%;  Pred. No. 0.8;
  Matches   91;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              || | ||| || |      ||  | ||||  ||     |||       |  | |||| | |
Db      10282 GGATCTGGAAAAGATAATGAATTTGGCCTTGGCCTTGATTCACCTCAGAATGAAATTGCC 10341

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
```

```
Db      10342 CAGAATGGGAACAACGACCAGTCCATCACAGAGTGTTCCATTGCCACCACAGCAGAGTTT 10401

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACC 399

Db      10402 TCTCATGACACGGATGCCACAGAGATCGACTCTCTGGATGGCTATGACCTGCAAGATGAA 10461

Qy        400 CGTGTTG 406

Db      10462 GATGATG 10468
```

RESULT 7
US-09-489-039A-3211/c
; Sequence 3211, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3211
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3211

Query Match 7.1%; Score 32.2; DB 4; Length 1122;
Best Local Similarity 61.2%; Pred. No. 0.49;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370

Db       1007 AGCACCGGACCGATAAACAGGCCGATCATGCCGAAGGCGATCAGGCCGCCAATGACGCCG 948

Qy        371 CGCTGGTTAATTATGATCAGCTGGT 395

Db        947 GTGAGGATAAGGATCATTGGCAGGT 923
```

RESULT 8
5486473-3/c
;Patent No. 5486473
; APPLICANT: FUKITA, HIROYUKI;YOSHIDA, IWAO;TAKAGI, MITSUO;
;MANABE, SADAO;FUKAI, KONOSUKE
; TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,049
; FILING DATE: 09-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 809,255
; FILING DATE: 18-DEC-1991
; APPLICATION NUMBER: 279,685
; FILING DATE: 05-DEC-1988
; APPLICATION NUMBER: 932,419
; FILING DATE: 19-NOV-1986
;SEQ ID NO:3:
; LENGTH: 1491
5486473-3

Query Match 6.8%; Score 31.2; DB 6; Length 1491;
Best Local Similarity 50.0%; Pred. No. 1.3;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237

Db        204 CAAGTCACTGACCGAAGCTAGGTAACAGTAACTGCGCACATCTGCGAGGTTGGCTGCTTC 145

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297

Db        144 CATGTTCATCATTTTGACATCAATGGTTGGCTTGTCTTTTGACATTATGGTCACACAACT 85

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333

Db         84 ATCGCCTTCCAGTACCAGATCAACCCATGTAGCTCC 49
```

RESULT 9
US-09-540-236-492/c
; Sequence 492, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 492
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-492

Query Match 6.8%; Score 31.2; DB 4; Length 2436;
Best Local Similarity 53.7%; Pred. No. 1.7;
Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

```
Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268

Db       1511 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 1452

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325

Db       1451 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 1392

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369

Db       1391 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 1348
```

RESULT 10
US-09-596-002-31/c
; Sequence 31, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 65792
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 31
; PUBLICATION INFORMATION:
US-09-596-002-31

Query Match 6.8%; Score 31.2; DB 4; Length 65792;
Best Local Similarity 53.7%; Pred. No. 11;

```
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              ||||||| |    | ||  |      | | |  | || || ||||| |||   | |||
Db       6819 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 6760

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |      | |  || ||||  |||| ||||||  ||||  |||| ||  || |      |
Db       6759 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 6700

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |   || ||  ||  |||| ||  ||||   |  |||
Db       6699 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 6656


RESULT 11
US-08-508-761B-1
; Sequence 1, Application US/08508761B
; Patent No. 6027920
;  GENERAL INFORMATION:
;    APPLICANT:  Joliff, Gwennael
;    APPLICANT:  Guyonvarch, Armel
;    APPLICANT:  Purification, Relano
;    APPLICANT:  Duchiron, Francis
;    APPLICANT:  Renaud, Michel
;    TITLE OF INVENTION:  System for Protein Expression and
;    TITLE OF INVENTION:  Secretion Especially in Corynebacteria
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jacobson, Price, Holman & Stern, PLLC
;      STREET:  400 Seventh St. N.W.
;      CITY:  Washington D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/508,761B
;      FILING DATE:  31-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09652
;      FILING DATE:  29-JUL-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09870
;      FILING DATE:  02-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Player, William E.
;      REGISTRATION NUMBER:  31,409
;      REFERENCE/DOCKET NUMBER:  P58525NA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 638-6666
;      TELEFAX:  (202) 393-5350
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2547 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Corynebacterium Melassecola
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  420..2390
US-08-508-761B-1

  Query Match             6.8%;  Score 30.8;  DB 3;  Length 2547;
  Best Local Similarity   58.9%;  Pred. No. 2.4;

  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
              |||| |||||  |  | | |  ||| ||| || ||||   ||||  | ||   ||  |||
Db        876 CAGAGTGGTTGGACCATTGAGACCAACATTGAGCAGTACTACGCCGATAAGAACGCCATT 935

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |||   || |  ||| ||  | | | ||
Db        936 GTTGTGCTCCCAATCGGTGGCGAGAGCTCC 965


RESULT 12
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4403765;
  Best Local Similarity   51.9%;  Pred. No. 95;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || ||||||| || ||  |  ||   | |||||| |||| |  |   || ||
Db    1357218 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357159

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
              ||     | | |    ||  | ||| | ||   |   ||   ||  || | |  |||
Db    1357158 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357099

Qy        165 TGCGCTTGCTCTG 177
              ||   |  |||
Db    1357098 CCCGGACGTGCTG 1357086


RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
```

```
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4411529;
  Best Local Similarity   51.9%;  Pred. No. 95;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || ||||||| || ||  |  ||   | |||||| |||| | |   || ||
Db    1357750 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357691

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
              ||    | | |    ||  | ||| | ||    |   ||   ||  || | |  |||
Db    1357690 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357631

Qy        165 TGCGCTTGCTCTG 177
              ||   |  |||
Db    1357630 CCCGGACGTGCTG 1357618


RESULT 14
US-09-350-756-2/c
; Sequence 2, Application US/09350756
; Patent No. 6495143
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Reseach Institute for Infectious  Diseases
;  APPLICANT: John S. Lee
;  APPLICANT: Peter Pushko
;  APPLICANT: Michael D. Parker
;  APPLICANT: Jonathan F. Smith
;  APPLICANT: Mark T. Dertzbaugh
;  APPLICANT: Leonard Smith
;  TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
;  FILE REFERENCE: 003/124/SAP RIID 98-21
;  CURRENT APPLICATION NUMBER: US/09/350,756
;  CURRENT FILING DATE: 1999-07-09
;  EARLIER APPLICATION NUMBER: US 60/092,416
;  EARLIER FILING DATE: 1998-07-10
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 2
;   LENGTH: 1341
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-2

  Query Match             6.6%;  Score 30;  DB 4;  Length 1341;
  Best Local Similarity   53.4%;  Pred. No. 3.2;
  Matches   63;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy        330 CTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCA 389
              || |||||||    |  | || | ||| | | |  | |   |   | || ||  ||| |
Db        673 CTTCGATATTGGACTGAGACAGTTCGGTGTTGAAGATGGAGAAGTATTTCATCCAGATGA 614

Qy        390 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAAC 447
              ||||  |||    | ||| |||    |  | | | |||    || ||  |  | ||||
Db        613 ACTGGGTACGATCGATGTCACCGTCCAGTTTGAAGATGATTTCACCGTTAGCGATAAC 556


RESULT 15
US-09-252-991A-10904
; Sequence 10904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10904
;   LENGTH: 1008
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904

  Query Match             6.5%;  Score 29.8;  DB 4;  Length 1008;
  Best Local Similarity   54.0%;  Pred. No. 3.2;
  Matches   61;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
              | | || | |  |  || | |  | ||| | |    ||||  ||||||  | |     |
Db        141 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 200

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              |||| |   | ||   || ||||| | |  ||||| || |     || ||| |
Db        201 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 253


Search completed: March 18, 2004, 20:20:26
Job time : 68.4971 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	34.2	7.5	616	12	US-10-424-599-88945	Sequence 88945, A
c	2	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c	3	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c	4	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c	5	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c	6	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
	7	34	7.5	2886	9	US-09-801-368-131	Sequence 131, App
	8	33.6	7.4	2751	15	US-10-398-221-3563	Sequence 3563, Ap
	9	33.6	7.4	3532	14	US-10-017-161-1913	Sequence 1913, Ap
	10	33.6	7.4	3532	15	US-10-292-798-1569	Sequence 1569, Ap
	11	33.4	7.3	7766	12	US-10-222-566-3	Sequence 3, Appli
	12	33.4	7.3	7766	14	US-10-222-162-3	Sequence 3, Appli
	13	33.4	7.3	7766	14	US-10-143-024-3	Sequence 3, Appli
	14	33.4	7.3	14770	10	US-09-873-367C-230	Sequence 230, App
	15	33.2	7.3	1355	15	US-10-045-674-593	Sequence 593, App
c	16	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
	17	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	18	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
	19	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c	20	32.2	7.1	1110	12	US-10-282-122A-23637	Sequence 23637, A
c	21	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
	22	32	7.0	1776	12	US-10-282-122A-24385	Sequence 24385, A
c	23	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
	24	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
	25	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
c	26	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
	27	31.4	6.9	840	12	US-10-282-122A-31900	Sequence 31900, A
	28	31.4	6.9	1245	15	US-10-369-493-40803	Sequence 40803, A
c	29	31.2	6.8	2394	12	US-10-282-122A-26837	Sequence 26837, A
c	30	31	6.8	525	10	US-09-918-995-32144	Sequence 32144, A
	31	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
	32	30.8	6.8	725	9	US-09-919-580-878	Sequence 878, App
	33	30.8	6.8	2050	14	US-10-086-510-1	Sequence 1, Appli
	34	30.8	6.8	2699	15	US-10-260-238-418	Sequence 418, App
	35	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
	36	30.6	6.7	2301	9	US-09-815-242-4844	Sequence 4844, Ap
	37	30.6	6.7	2301	9	US-09-815-242-9042	Sequence 9042, Ap
c	38	30.6	6.7	75216	14	US-10-080-170-646	Sequence 646, App
c	39	30.4	6.7	392	12	US-10-282-122A-25231	Sequence 25231, A
c	40	30.4	6.7	992	12	US-10-425-114-31747	Sequence 31747, A
c	41	30.4	6.7	1024	15	US-10-323-069A-111	Sequence 111, App
c	42	30.4	6.7	1601042	15	US-10-027-632-59064	Sequence 59064, A
	43	30.2	6.6	654	12	US-10-425-114-28546	Sequence 28546, A
c	44	30.2	6.6	829	15	US-10-027-632-164999	Sequence 164999,
c	45	30.2	6.6	1203	9	US-09-738-626-3184	Sequence 3184, Ap

ALIGNMENTS

RESULT 1
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(616)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

Query Match 7.5%; Score 34.2; DB 12; Length 616;
Best Local Similarity 50.9%; Pred. No. 0.51;
Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
              | | |  |||  ||  || ||| || || ||| | |    | |  | | | | |  | |
Db        556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497

Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              | | ||    || | | |   || ||  |  |   | | |||   ||  |  | | |
Db        496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437
```

```
Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              || ||||||  |  |   ||| |||  ||||| ||| ||
Db        436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTA 398
```

RESULT 2
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              ||||  || || ||   | ||  | || ||||   || || |   ||||  |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||     |     ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 3
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              ||||  || || ||   | ||  | || ||||   || || |   ||||  |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||     |     ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 4
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              ||||  || || ||   | ||  | || ||||   || || |   ||||  |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||     |     ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 5
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora

US-10-010-390-6

Query Match 7.5%; Score 34; DB 14; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 6
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (EBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-441-736-5

Query Match 7.5%; Score 34; DB 15; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 7
US-09-801-368-131
; Sequence 131, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 131
; LENGTH: 2886
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-131

Query Match 7.5%; Score 34; DB 9; Length 2886;
Best Local Similarity 50.6%; Pred. No. 1.4;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              | ||  | | | | || | | | |||||    | ||    || |   |    |||| | |
Db       1501 AACAACAACAATAACGATAACGATAACGATAACAATAATAGCAATAATAATAATAACAAT 1560

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || | |    | | | |||  |  ||  ||||||  ||  | | ||  |    |||  |
Db       1561 AATAATAATAACAATAATAATAATAACAATAATAATAACAATAACAACAACAACAACAAT 1620

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
              || ||     ||| || |        || |   || | |||||
Db       1621 AATAATAATAATAGTAATAATATTAAAAATAACGTTGATAAC 1662
```

RESULT 8
US-10-398-221-3563
; Sequence 3563, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3563
; LENGTH: 2751
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3563

Query Match 7.4%; Score 33.6; DB 15; Length 2751;
Best Local Similarity 51.3%; Pred. No. 1.9;
Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||||||   | |  ||| | |||| ||||||  ||      |       | || | | ||
Db        321 GATGTAGTGAAAGAAGCGAAAAATATTACTATCGACAAACCGTTTCCTCGTATGACATAT 380
```

```
Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              || |   | || || |  |     | | ||  | ||||||  | | ||   | |
Db        381 AAAGAAGCAATGGATCGTTTTGGCAGTGATAAACCAGATATTCGTTTTGGTTTAGAATTA 440

Qy        358 GGTAATAACGCCGCGCTGGTTAATTATGATCA 389
                |||   |||   | || ||  ||| | |
Db        441 CAAAATGTATCCGATGTCGTAAAAGATGTTGA 472


RESULT 9
US-10-017-161-1913
; Sequence 1913, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1913
;   LENGTH: 3532
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(3532)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(437)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2622)..(3332)
US-10-017-161-1913

  Query Match             7.4%;  Score 33.6;  DB 14;  Length 3532;
  Best Local Similarity   46.6%;  Pred. No. 2.2;
  Matches  108;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy         215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
               | |  ||| ||||||| ||      |||  |  | ||||  | | ||||||  | ||
Db        2881 ATAATGGTGATGGTAATGGTAATCATGGTGATGATGGTGGTGGTGATAGTAATCATGGTG 2940

Qy         275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
               | |  ||| |    |  | |||                |  |    |  |  |    |
Db        2941 ACGGTAATGATGATAGTGGTGGTGATGATGATGATGGTGATGATGATGGAGATCATGTTG 3000

Qy         335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGG 394
               || ||  ||| |      |  |||| || | |  |      | ||||  ||||||    |
Db        3001 ATGTTGATGTTGATAGTGATAACGGTGATGATGGTGGTGATGATAATGGTGATCACGGTG 3060

Qy         395 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAA 446
               ||    ||| ||||    ||| |  |  ||  ||  | | ||  | | |
Db        3061 TTGATGGTGGTGTTGATGGTGATAATGGTGATGGTAATGATGGCAATGGTGA 3112


RESULT 10
US-10-292-798-1569
; Sequence 1569, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1569
;   LENGTH: 3532
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(3532)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(437)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2622)..(3332)
US-10-292-798-1569

  Query Match             7.4%;  Score 33.6;  DB 15;  Length 3532;
  Best Local Similarity   46.6%;  Pred. No. 2.2;
  Matches  108;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy         215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
               | |  ||| ||||||| ||      |||  |  | ||||  | | ||||||  | ||
Db        2881 ATAATGGTGATGGTAATGGTAATCATGGTGATGATGGTGGTGGTGATAGTAATCATGGTG 2940

Qy         275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
               | |  ||| |    |  | |||                |  |    |  |  |    |
Db        2941 ACGGTAATGATGATAGTGGTGGTGATGATGATGATGGTGATGATGATGGAGATCATGTTG 3000

Qy         335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGG 394
               || ||  ||| |      |  |||| || | |  |      | ||||  ||||||    |
Db        3001 ATGTTGATGTTGATAGTGATAACGGTGATGATGGTGGTGATGATAATGGTGATCACGGTG 3060

Qy         395 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAA 446
               ||    ||| ||||    ||| |  |  ||  ||  | | ||  | | |
Db        3061 TTGATGGTGGTGTTGATGGTGATAATGGTGATGGTAATGATGGCAATGGTGA 3112


RESULT 11
US-10-222-566-3
; Sequence 3, Application US/10222566
; Publication No. US20040044192A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD3
;  CURRENT APPLICATION NUMBER: US/10/222,566
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-566-3
```

```
  Query Match             7.3%;  Score 33.4;  DB 12;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 4;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||     ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |   |   | ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |     | ||| |          | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||        | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 12
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD4
;  CURRENT APPLICATION NUMBER: US/10/222,162
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-162-3

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 4;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||     ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |   |   | ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |     | ||| |          | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||        | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 13
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD1
;  CURRENT APPLICATION NUMBER: US/10/143,024
;  CURRENT FILING DATE:  2002-05-10
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-143-024-3

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 4;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||     ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |   |   | ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |     | ||| |          | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||        | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 14
US-09-873-367C-230
; Sequence 230, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  APPLICANT:  Soppet, Daniel
;  APPLICANT:  Endress, Gregory
;  APPLICANT:  Augustus, Meena
;  APPLICANT:  Ebner, Reinhard
;  APPLICANT:  Carter, Kenneth
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
;  TITLE OF INVENTION:  Signature Gene Sets
;  FILE REFERENCE: 689290-64
;  CURRENT APPLICATION NUMBER: US/09/873,367C
;  CURRENT FILING DATE:  2003-04-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,891
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,842
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/244,867
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: U.S. 60/245,084
;  PRIOR FILING DATE: 2000-11-01
```

```
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 230
;   LENGTH: 14770
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-873-367C-230

  Query Match             7.3%;  Score 33.4;  DB 10;  Length 14770;
  Best Local Similarity   48.7%;  Pred. No. 5.7;
  Matches   91;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              || | ||| || |      ||  | ||||  ||     |||      |   | |||| | |
Db      10282 GGATCTGGAAAAGATAATGAATTTGGCCTTGGCCTTGATTCACCTCAGAATGAAATTGCC 10341

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
              ||||||||   ||   |  |  ||| ||  ||   |  |  || |   |   | ||  ||
Db      10342 CAGAATGGGAACAACGACCAGTCCATCACAGAGTGTTCCATTGCCACCACAGCAGAGTTT 10401

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACC 399
               ||   | |    | | |   | | || | | |||| |   ||||| | ||  | |
Db      10402 TCTCATGACACGGATGCCACAGAGATCGACTCTCTGGATGGCTATGACCTGCAAGATGAA 10461

Qy        400 CGTGTTG 406
              || ||
Db      10462 GATGATG 10468


RESULT 15
US-10-045-674-593
; Sequence 593, Application US/10045674
; Publication No. US20030232333A1
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT: HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 60/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674-593

  Query Match             7.3%;  Score 33.2;  DB 15;  Length 1355;
  Best Local Similarity   23.9%;  Pred. No. 1.8;
  Matches   83;  Conservative   91;  Mismatches  174;  Indels    0;  Gaps    0;

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137
              :||:||:||: ::  :  :  : ::||:||:  ::::   ||:   | :::   :   |
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197
              :  :: :  :  :||::::| :: :| :|   |:| :    |:|    ||   :
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGA 257
              : |:  : |:|| ||: |:| :| : |:|  ||:|      |  | :      |   |
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        258 TAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              :  :::: | |       : |   || |  :|   |:  :  :| :  :||:   ::
Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYATYGGYGAYGTYWS 1007

Qy        318 GAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGT 377
              :  :  ||:  : |:  : |:  :||  | | :|  ||:  :||  |:  : |||
Db       1008 YGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKAATTCYCARATGGC 1067

Qy        378 TAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 425
              : |  :| : |   :| :| : ::  : |:|   |: | |   :||
Db       1068 YCARGTYGGWGAYGGKGAYAAYWSWCCKYTWATGAAYAAYTTYMGWCA 1115


Search completed: March 19, 2004, 03:49:15
Job time : 257.734 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-11	Sequence 11, Appl
2	404.8	88.8	456	6	US-08-233-642A-56	Sequence 56, Appl
3	404.8	88.8	456	23	US-09-543-407-1	Sequence 1, Appli
4	376	82.5	456	23	US-09-543-407-13	Sequence 13, Appl
5	361.6	79.3	456	23	US-09-543-407-19	Sequence 19, Appl
6	355.2	77.9	456	23	US-09-543-407-17	Sequence 17, Appl
7	353.6	77.5	456	23	US-09-543-407-23	Sequence 23, Appl
8	353.6	77.5	456	23	US-09-543-407-25	Sequence 25, Appl
9	352	77.2	456	23	US-09-543-407-15	Sequence 15, Appl
10	350.4	76.8	456	23	US-09-543-407-21	Sequence 21, Appl
11	350.4	76.8	456	23	US-09-543-407-27	Sequence 27, Appl
12	347.2	76.1	456	23	US-09-543-407-29	Sequence 29, Appl
13	315.2	69.1	361	6	US-08-233-642A-54	Sequence 54, Appl
14	222.4	48.8	477	17	US-09-252-691-172	Sequence 172, App
15	222.4	48.8	477	17	US-09-252-691C-172	Sequence 172, App
16	222.4	48.8	477	51	US-10-417-886-172	Sequence 172, App
17	216	47.4	456	23	US-09-543-407-3	Sequence 3, Appli
18	216	47.4	675	13	US-08-978-878-3	Sequence 3, Appli
19	216	47.4	675	31	US-09-741-873B-3	Sequence 3, Appli
20	213.6	46.8	453	92	US-60-352-946-1	Sequence 1, Appli
21	213.6	46.8	453	101	US-60-444-371-1	Sequence 1, Appli
22	172	37.7	396	13	US-08-978-878-1	Sequence 1, Appli
23	163.2	35.8	360	31	US-09-741-873B-1	Sequence 1, Appli
24	52.2	11.4	78	23	US-09-543-407-53	Sequence 53, Appl
c 25	49.4	10.8	78	23	US-09-543-407-52	Sequence 52, Appl
26	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
c 27	40.8	8.9	78	23	US-09-543-407-48	Sequence 48, Appl
28	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
29	40.6	8.9	78	23	US-09-543-407-49	Sequence 49, Appl
c 30	39.4	8.6	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
31	39	8.6	100	45	US-10-146-492B-78	Sequence 78, Appl
32	37.4	8.2	561	17	US-09-252-691-171	Sequence 171, App
33	37.4	8.2	561	17	US-09-252-691C-171	Sequence 171, App
34	37.4	8.2	561	51	US-10-417-886-171	Sequence 171, App
35	37.4	8.2	1304	51	US-10-437-963-19773	Sequence 19773, A
36	35.4	7.8	100	45	US-10-146-492B-77	Sequence 77, Appl
c 37	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 24330, A
c 38	35.4	7.8	3412	1	PCT-US01-08631-24948	Sequence 24948, A
39	34.6	7.6	456	23	US-09-543-407-4	Sequence 4, Appli
c 40	34.2	7.5	385	18	US-09-333-535A-288	Sequence 288, App
c 41	34.2	7.5	385	25	US-09-615-606A-909	Sequence 909, App
c 42	34.2	7.5	396	64	US-60-089-524-4522	Sequence 4522, Ap
c 43	34.2	7.5	616	51	US-10-424-599-88945	Sequence 88945, A
c 44	34.2	7.5	880	27	US-09-654-617-2638	Sequence 2638, Ap
c 45	34.2	7.5	880	29	US-09-684-016-2638	Sequence 2638, Ap

ALIGNMENTS

```
RESULT 1
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 7.6e-136;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 2.6e-119;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||| |           | | || | ||   |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 2.6e-119;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||| |           | | || | ||   |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             82.5%;  Score 376;  DB 23;  Length 456;
  Best Local Similarity   89.0%;  Pred. No. 5.1e-110;
  Matches  406;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.3e-105;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.6e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 8.6e-103;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              || |    || |  | ||| |          | | || | ||  |    | | | |
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 8.6e-103;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.8e-102;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |  ||  ||   | | | |  |  | |    |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 9.3e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||||  |  |        | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 9.3e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                  |  |||   |     |   |  ||   .|     |     |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                ||||||||||||||||||||  |           | | || | ||   |     | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
                |    ||  ||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

  Query Match             76.1%;  Score 347.2;  DB 23;  Length 456;
  Best Local Similarity   85.1%;  Pred. No. 1e-100;
  Matches  388;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
                ||||||   || |||  |    |                  |   | | ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                ||||||||||||||||||||  |           | | || | ||   |     | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
                |    ||  ||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-642A-54

  Query Match             69.1%;  Score 315.2;  DB 6;  Length 361;
  Best Local Similarity   99.1%;  Pred. No. 2e-90;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
                ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              ||||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320


RESULT 14
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match             48.8%;  Score 222.4;  DB 17;  Length 477;
  Best Local Similarity   68.0%;  Pred. No. 1.9e-60;
  Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| ||||||||||    | ||||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||      |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||||| |||||||||| ||||||||   || |||| | |||||||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||    || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||    ||| |||| | ||| |  ||     || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  |||| |||  |  |      |  |  |   ||  |  ||  | |  |  |
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||  ||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


RESULT 15
US-09-252-691C-172
; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-172

  Query Match             48.8%;  Score 222.4;  DB 17;  Length 477;
  Best Local Similarity   68.0%;  Pred. No. 1.9e-60;
  Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| ||||||||||    | ||||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||      |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||||| |||||||||| ||||||||   || |||| | |||||||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||    || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||    ||| |||| | ||| |  ||     || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  |||| |||  |  |      |  |  |   ||  |  ||  | |  |  |
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||  ||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


Search completed: March 19, 2004, 03:01:49
Job time : 2651.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	216	47.4	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	172	37.7	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33.4	7.3	3990	5	US-09-830-230A-629	Sequence 629, App
	4	33.2	7.3	1355	6	US-10-045-674A-593	Sequence 593, App
	5	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
	6	31.4	6.9	1588	6	US-10-767-701-14522	Sequence 14522, A
c	7	31.2	6.8	290892	6	US-10-767-471-10809	Sequence 10809, A
c	8	30.2	6.6	512	6	US-10-781-014-257	Sequence 257, App
c	9	30.2	6.6	1326	6	US-10-781-014-255	Sequence 255, App
c	10	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
	11	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
c	12	29.4	6.4	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	13	29.4	6.4	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	14	29.4	6.4	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	15	29.4	6.4	1400	7	US-60-545-213-6407	Sequence 6407, Ap
	16	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	17	28.4	6.2	1350	6	US-10-784-480-291	Sequence 291, App
	18	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	19	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	20	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	21	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
	22	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
c	23	28	6.1	600	7	US-60-545-213-1313	Sequence 1313, Ap
c	24	28	6.1	600	7	US-60-545-213-1314	Sequence 1314, Ap
c	25	28	6.1	600	7	US-60-545-213-5585	Sequence 5585, Ap
c	26	28	6.1	600	7	US-60-545-213-5586	Sequence 5586, Ap
c	27	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
c	28	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
	29	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
	30	27.8	6.1	1580	6	US-10-451-467A-239	Sequence 239, App
c	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	32	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	33	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	34	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	35	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	36	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	37	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
c	38	27.4	6.0	1295	6	US-10-100-683-5480	Sequence 5480, Ap
	39	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
c	40	27.4	6.0	3416	6	US-10-100-683-3397	Sequence 3397, Ap
c	41	27.4	6.0	8206	6	US-10-021-698A-3529	Sequence 3529, Ap
	42	27.4	6.0	186449	6	US-10-021-698A-715	Sequence 715, App
	43	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
	44	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
	45	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 47.4%; Score 216; DB 5; Length 675;
Best Local Similarity 67.1%; Pred. No. 5.9e-62;
Matches 306; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||    ||     |||||  || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                ||||  ||    ||| |  |  |           ||   ||          | | |
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 37.7%; Score 172; DB 5; Length 396;
Best Local Similarity 64.6%; Pred. No. 1.8e-47;
Matches 256; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||     || |||||  |||| |  ||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||    ||     |||||  || || |||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
```

```
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || ||    || | ||| |||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                ||||  ||    ||| |  |  |           ||   ||          | | |
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
; LENGTH: 3990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1139)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1148)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1210)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1244)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1247)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1250)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1251)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

```
  Query Match             7.3%;  Score 33.4;  DB 5;  Length 3990;
  Best Local Similarity   47.4%;  Pred. No. 0.38;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              |  ||  ||||||  ||||    |    ||||      ||  ||| |  | | |  |  | |
Db       3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |    |  ||||||||||  |  |||    |||  ||
Db       3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    |  ||| |           | |||  |||||  |     |
Db       3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 3612

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
               || ||||       | |  ||| | || |
Db       3613 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 3643


RESULT 4
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674A
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 06/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674A-593

  Query Match             7.3%;  Score 33.2;  DB 6;  Length 1355;
  Best Local Similarity   23.9%;  Pred. No. 0.29;
  Matches   83;  Conservative   91;  Mismatches  174;  Indels    0;  Gaps    0;

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137
              :||:||:||: ::  :  :  :  ::||:||:  ::::   ||:  | :::  :    |
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197
              :  :: :  :  :||::::| :: :| :|   |:| :    |:|      ||   :
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGA 257
              : |:  :  |:|| ||: |:| :| : |:|  ||:|      |    | :      |   |
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        258 TAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              :  :::: | |        : |   || | :|   |:  :  :| :  :||:   ::
Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYATYGGYGAYGTYWS 1007

Qy        318 GAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGT 377
              :  :   ||: : |:  : |:  :||  | | :|  ||:  :||  |:  : |||
Db       1008 YGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKAATTCYCARATGGC 1067

Qy        378 TAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 425
              : |   :| : |   :| :| : ::  : |:|   |: |  |   :||
Db       1068 YCARGTYGGWGAYGGKGAYAAYWSWCCKYTWATGAAYAAYTTYMGWCA 1115


RESULT 5
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
;  APPLICANT: Neubauer, Antonie
;  APPLICANT:  Ziegler, Christina
;  TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
;  FILE REFERENCE: 1/1372
;  CURRENT APPLICATION NUMBER: US/10/624,149A
;  CURRENT FILING DATE:  2003-07-21
;  PRIOR APPLICATION NUMBER: 60/403,282
;  PRIOR FILING DATE: 2002-08-14
;  PRIOR APPLICATION NUMBER: DE 10233064
;  PRIOR FILING DATE: 2002-07-19
;  PRIOR APPLICATION NUMBER: DE 10317008
;  PRIOR FILING DATE: 2003-04-11
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
;   LENGTH: 150223
;   TYPE: DNA
;   ORGANISM: Equine herpesvirus 1
;  PUBLICATION INFORMATION:
;   AUTHORS: Telford,E.A.
;   AUTHORS: Watson, M.S.
;   AUTHORS: McBride, K.
;   AUTHORS: Davison, A.J.
;   TITLE: The DNA sequence of equine herpesvirus-1
;   JOURNAL: Virology
;   VOLUME: 189
;   ISSUE: 1
;   PAGES: 304-316
;   DATE: JUL-1992
;   DATABASE ACCESSION NUMBER: NC 001491, NCBI
;   DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 150223;
  Best Local Similarity   51.4%;  Pred. No. 3.6;
  Matches   75;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              || ||  ||  |  | || ||  | ||| ||| |   |  ||| |||   | ||   |||
Db      95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy        233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
              |     |  |||| ||  ||||||    | || |  | | | ||| |     |  |||
Db      95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy        293 GAAATAATGCCACCATCGACCAGTGG 318
              || |   |     || ||| | || |
Db      95662 GAGAGCTTATTCACAACGAGCTGTTG 95687


RESULT 6
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
```

```
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
;   LENGTH: 1588
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 1588;
  Best Local Similarity   44.5%;  Pred. No. 1.2;
  Matches  125;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
Db        455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
Db        515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy        160 AACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
Db        575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
Db        635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAA 320
Db        695 GCTGGTAGCTTTGGCGGTGGTTCCACTGGCGCCGAGAAGGA 735


RESULT 7
US-10-767-471-10809/c
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
;   LENGTH: 290892
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(290892)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 290892;
  Best Local Similarity   44.6%;  Pred. No. 12;
  Matches  123;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db     234133 GATTATGATGATGATGATGGTAATGGTGATGATGGTGATGGTGATAGTTACAGTAGTGAT 234074

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db     234073 GTTGGTAATGATAGTGGTGGTGACGACGGTGATGCTAATGATTATGATGATGATGGTGGT 234014

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db     234013 GGTAATGGTGATGATGGTAATGTTGGTGATGATCGTGGCTGTGATGGTGATGGTTATGAT 233954

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGGTT 396
Db     233953 GGGAATGTTGGTGAAGATCGTGGCTGTGATGGTGATGGTGGTGATGATGATGACAATGAT 233894

Qy        397 ACCCGTGTTGTTACCCATGAAATGGCACATGCAAAC 432
Db     233893 AAGGATGGTAATGAAAATGATAATGCTGATGCTAAC 233858


RESULT 8
US-10-781-014-257/c
; Sequence 257, Application US/10781014
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
;  TITLE OF INVENTION:  INVOLVED IN CARBON METABOLISM AND ENERGY
;  TITLE OF INVENTION:  PRODUCTION
;  FILE REFERENCE: BGI-126CPCN
;  CURRENT APPLICATION NUMBER: US/10/781,014
;  CURRENT FILING DATE:  2004-02-17
;  PRIOR APPLICATION NUMBER: US 09/602,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/143,208
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: 60/151,572
;  PRIOR FILING DATE: 1999-08-31
;  PRIOR APPLICATION NUMBER: DE 19931412.8
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931413.6
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931424.1
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931428.4
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 257
;   LENGTH: 512
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(489)
;   OTHER INFORMATION: FRXA02596
US-10-781-014-257

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 512;
  Best Local Similarity   56.6%;  Pred. No. 1.9;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        310 GACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
Db        375 GCCCAGTCGACCGCCGAACAGCACCTTATTATTAGCAGCCTCTTCAGCAGCCAGAAGGCG 316

Qy        370 GCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTT 408
Db        315 GTACTGCTTCAGCATGTCTCGGTCGTCTGGAGTGTTGAT 277
```

```
RESULT 9
US-10-781-014-255/c
; Sequence 255, Application US/10781014
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
;  TITLE OF INVENTION:  INVOLVED IN CARBON METABOLISM AND ENERGY
;  TITLE OF INVENTION:  PRODUCTION
;  FILE REFERENCE: BGI-126CPCN
;  CURRENT APPLICATION NUMBER: US/10/781,014
;  CURRENT FILING DATE:  2004-02-17
;  PRIOR APPLICATION NUMBER: US 09/602,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/143,208
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: 60/151,572
;  PRIOR FILING DATE: 1999-08-31
;  PRIOR APPLICATION NUMBER: DE 19931412.8
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931413.6
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931424.1
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931428.4
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 255
;   LENGTH: 1326
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101)..(1303)
;   OTHER INFORMATION: RXN02596
US-10-781-014-255

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 1326;
  Best Local Similarity   56.6%;  Pred. No. 2.8;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        310 GACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              | ||||| || ||| || |   |  |||||  |  | |||  | | || |  | || ||
Db       1189 GCCCAGTCGACCGCCGAACAGCACCTTATTATTAGCAGCCTCTTCAGCAGCCAGAAGGCG 1130

Qy        370 GCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTT 408
              |  ||| || |  |||    | | ||     ||||||
Db       1129 GTACTGCTTCAGCATGTCTCGGTCGTCTGGAGTGTTGAT 1091


RESULT 10
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
;   LENGTH: 24900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(24900)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-767-471-10736

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 24900;
  Best Local Similarity   52.4%;  Pred. No. 15;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||  | || ||  | || | ||  |       |   | || || |  | | ||||| |
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              | |||||  |  | | |||||  | ||||| ||| | | | |     || |   |||
Db       9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929

Qy        327 AAAC 330
              ||||
Db       9928 AATC 9925


RESULT 11
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 2.8;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||    |      |||| | |     | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
                |||||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 12
US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
```

```
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2134
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2134

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 5.3;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGC 391
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAG 450
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 13
US-60-545-213-2135/c
; Sequence 2135, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2135
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2135

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 5.3;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGC 391
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAG 450
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 14
US-60-545-213-6406/c
; Sequence 6406, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6406
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6406

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 5.3;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGC 391
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAG 450
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 15
US-60-545-213-6407/c
; Sequence 6407, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6407
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6407

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 5.3;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGC 391
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAG 450
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199
```

Search completed: March 19, 2004, 03:10:07
Job time : 55.1445 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	216.6	47.5	680	12	BJ618688	BJ618688 BJ618688
	2	35.2	7.7	813	28	BZ204853	BZ204853 CH230-390
	3	34.8	7.6	801	12	BI648266	BI648266 603278209
	4	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
	5	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	6	34	7.5	234	10	BB581987	BB581987 BB581987
c	7	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	8	34	7.5	811	28	BZ469000	BZ469000 BOOAO58TF
	9	33.8	7.4	296	9	AV058630	AV058630 AV058630
c	10	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	11	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	12	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
c	13	33.8	7.4	695	28	AQ637195	AQ637195 RPCI-11-4
	14	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	15	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	16	33.6	7.4	374	13	BY236219	BY236219 BY236219
	17	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	18	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
c	19	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	20	33.4	7.3	505	6	AL809250	Al809250 Triticum
	21	33.4	7.3	624	14	CB578319	CB578319 AMGNNUC:N
	22	33.4	7.3	664	13	BX506928	BX506928 DKFZp779I
	23	33.4	7.3	679	29	CC961526	CC961526 BOIGS10TF
c	24	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
c	25	33.4	7.3	699	12	BI959328	BI959328 HVSMEn001
c	26	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF
c	27	33.4	7.3	709	28	BH690914	BH690914 BOHXR64TF
	28	33.4	7.3	715	13	BU444683	BU444683 603765715
c	29	33.4	7.3	785	28	BH578793	BH578793 BOGLY20TR
c	30	33.4	7.3	815	28	BH461716	BH461716 BOHIV03TR
c	31	33.4	7.3	827	28	BH438448	BH438448 BOGVQ28TR
c	32	33.4	7.3	830	28	BH715216	BH715216 BOMIM42TR
c	33	33.4	7.3	835	28	BH575194	BH575194 BOHJO27TF
c	34	33.4	7.3	853	28	BH578402	BH578402 BOHNV61TF
	35	33.4	7.3	856	14	CD456845	CD456845 Fg05_03n1
c	36	33.4	7.3	1026	28	BZ458784	BZ458784 BONHB14TF
	37	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
	38	33.2	7.3	592	29	CG072975	CG072975 PUICN43TB
	39	33.2	7.3	629	9	AU295325	AU295325 AU295325
	40	33.2	7.3	675	28	AQ361497	AQ361497 mgxb0004K
	41	33.2	7.3	788	9	AU139321	AU139321 AU139321
c	42	33.2	7.3	1646	10	BF579120	BF579120 602096054
	43	33	7.2	456	12	BJ331339	BJ331339 BJ331339
	44	33	7.2	524	12	BJ367372	BJ367372 BJ367372
	45	33	7.2	547	12	BJ333958	BJ333958 BJ333958

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 47.5%; Score 216.6; DB 12; Length 680;
Best Local Similarity 67.1%; Pred. No. 8e-51;
Matches 306; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||  ||    ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||   ||| |  |   |          ||  ||          | | |
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
BZ204853
LOCUS BZ204853 813 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-390D21, genomic survey sequence.
ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 813)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-390D21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 390 row: D column: 21
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .813
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"
ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 813;
Best Local Similarity 49.0%; Pred. No. 59;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```
Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||  |   | | | | || |   |   ||  ||  | ||  |  |  ||  |  ||
Db        548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              || |   || ||       || |  | |  |||||  |     |||||  ||  |   |
Db        608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              ||  ||  |||||||| ||| |   | ||||| | |   | |||| ||| ||      |
Db        668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAAGGACAC 727

Qy        358 GGTAATAACGCC 369
              ||  |.| | |
Db        728 TGTGGTTAGGAC 739
```

RESULT 3
BI648266
LOCUS BI648266 801 bp mRNA linear EST 12-SEP-2001
DEFINITION 603278209F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5318307 5', mRNA sequence.
ACCESSION BI648266
VERSION BI648266.1 GI:15562502
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 801)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

```
            http://image.llnl.gov
            Plate: LLAM11807  row: b  column: 04
            High quality sequence stop: 798.
FEATURES              Location/Qualifiers
     source           1. .801
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="129,C57BL/6J,FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5318307"
                      /tissue_type="tumor, gross tissue"
                      /dev_stage="10 months"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_Mam3"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigators
                      providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                      Reference for transgenic model: Xu et al., Nature Genetics
                      22, 37-43 (1999)."
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 12;  Length 801;
  Best Local Similarity   50.0%;  Pred. No. 76;
  Matches  114;  Conservative    0;  Mismatches  112;  Indels    2;  Gaps    1;

Qy        210 TACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTAT 269
              | | |  |||    ||||| || ||  |          | ||  | ||  ||  | ||
Db        322 TTCACCCAGCAAGCATGGTGACCGCTTCATCCCCTCGCGGGCCGGGGCCAACTGGACGCG 381

Qy        270 TGAACTGACTCAGAATGGTTTCAGAA--ATAATGCCACCATCGACCAGTGGAACGCTAAA 327
              ||||||  | ||| ||    |  | ||  | ||  || ||| | |  |  | || || ||
Db        382 TGAACTTCCACAGGATCAATGAACAATGAGAAGTCCCCCAGCCAGAACCGCAAAGCCAAG 441

Qy        328 AACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGAT 387
               || ||          ||| || |||||       || |||| ||| | ||  ||||
Db        442 GACGCCACCTCGGACAATGGCAAAGACGGCCTGGCCTACTCCGCACTGCTGAAGAATGAG 501

Qy        388 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAAC 435
              | ||||| | || |  | |  |    | |     ||||  |  |  ||
Db        502 CTGCTGGGTGCCGGCATCGAGAAGGTTCAGGACCCACAGACGGAGGAC 549


RESULT 4
BJ334624
LOCUS       BJ334624                523 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda47o06 5', mRNA sequence.
ACCESSION   BJ334624
VERSION     BJ334624.1  GI:19164754
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .523
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="dda47o06"
                      /sex="mat A"
                      /dev_stage="Aggregation stage"
                      /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 523;
  Best Local Similarity   57.1%;  Pred. No. 93;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        398 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGANGACAATAAT 457

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACT 342
              |||  || || | || |    ||   ||||||     |||||| |
Db        458 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAAT 502


RESULT 5
BG671547
LOCUS       BG671547                1036 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5',
            mRNA sequence.
ACCESSION   BG671547
VERSION     BG671547.1  GI:13893646
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C.,
            Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D.,
            Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
   PUBMED   12060780
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .1036
                      /organism="Rattus norvegicus"
                      /mol_type="mRNA"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="DRNBUB05"
                      /sex="male"
                      /tissue_type="dorsal root ganglion"
                      /dev_stage="adult"
                      /clone_lib="Rat DRG Library"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 1036;
  Best Local Similarity   49.2%;  Pred. No. 1.3e+02;
  Matches   90;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;
```

```
Qy          11 TAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTC 70
               |||||   ||  | |  |||    |||  || | | || |||| |     |||| |
Db         168 TGAAAGATACATGAGTGACAGTGTCCGTGATTGCAGACATTGCTGTATGATGCGTCCTGT 227

Qy          71 CACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAA 130
               |   | | | |  ||  |||||  ||  ||  | |   || | |   |   ||     |
Db         228 CTGTAGGAGTCCTCGCTGGTAACGATGCCGAGAGTGCACATCGATATCGGTGGGCGGTTA 287

Qy         131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
               | |     |  |  ||   | |    | ||| |||   ||| ||||  ||| || ||
Db         288 CATCTGAAGTGGAAGAGGTGGCTCTTTCGAACTCTGACTTTGATCTGTCAAGGGACGCTT 347

Qy         191 GTA 193
               |||
Db         348 GTA 350
```

RESULT 6
BB581987/c
LOCUS BB581987 234 bp mRNA linear EST 30-NOV-2000
DEFINITION BB581987 RIKEN full-length enriched, adult male colon Mus musculus cDNA clone 9030005O05 5', mRNA sequence.
ACCESSION BB581987
VERSION BB581987.1 GI:11478531
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 234)
AUTHORS Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .234
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9030005O05"
/sex="male"
/tissue_type="colon"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male colon"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
ORIGIN

Query Match 7.5%; Score 34; DB 10; Length 234;
Best Local Similarity 51.3%; Pred. No. 73;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy          35 TCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATC 94
               | ||  ||  | |    | |  | || ||  || |    ||   || | || || ||   |
Db         200 TAGTTTTTGGTAGGTCGGTTTATGCGGGTTTCTTCTTCCACCAGGCCCGCCGCCGTTTCC 141

Qy          95 ATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTT 154
               ||||  |||  ||  ||   ||      |   |  || | ||||||| ||  ||  ||
Db         140 CGAACGCCGGGTGCTTTATAGCCTTGTTCAAAATATGGTAGCGCATTTACCACCACCTTT 81

Qy         155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
               ||| | ||   || | ||| |  ||  |||  ||
Db          80 CCGGTTACCAAGCCCGTGCCCGCCATTGCGCGGC 47
```

RESULT 7
CA711477/c
LOCUS CA711477 500 bp mRNA linear EST 26-NOV-2002
DEFINITION wdk2c.pk014.o3 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.o3 5' end, mRNA sequence.
ACCESSION CA711477
VERSION CA711477.1 GI:25433270
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 500)
AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE DuPont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES Location/Qualifiers
source 1. .500
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"

```
                     /clone="wdk2c.pk014.o3"
                     /tissue_type="kernel"
                     /clone_lib="wdk2c"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
                     days after anthesis."
ORIGIN

  Query Match             7.5%;  Score 34;  DB 14;  Length 500;
  Best Local Similarity   54.5%;  Pred. No. 1e+02;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  |  ||| |||||||   ||   |  | ||| ||      | | ||  | ||| |
Db        274 AGNTGNTGCAGTCGCAGCCNCCGGTCTCCCCGGCCGTNTCGAACCCGCCGGAGCTCCCCN 215

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGAC 126
                 ||| ||||| ||  || |  | |||||||||     |    ||||   ||
Db        214 TGTGGGAGGCGGTGGGCATGACCATGGCGGCGGCGCCGGCGGTGGCCNCCAC 163


RESULT 8
BZ469000/c
LOCUS       BZ469000                811 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BOOAO58TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOAO58,
            genomic survey sequence.
ACCESSION   BZ469000
VERSION     BZ469000.1  GI:26764546
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOOAO58TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOOAO58"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 28;  Length 811;
  Best Local Similarity   48.9%;  Pred. No. 1.3e+02;
  Matches   91;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        176 TGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCG 235
              ||  | || |     ||||| |||       ||||| |     |  | |||||  ||  ||
Db        610 TGTTAGGCAAGCATTGTAAAACTGTGTTTCCCATTTCATCTACTGTTTATGACAAATGCT 551

Qy        236 CCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
                ||| ||   ||   ||  | |   | | ||  |    || ||  | || |||| | |
Db        550 TTGATCTAATTCACTCTGATGTTTGGACTGCTCCTTGCTTGTCTAGGGATAGTTTTAAAT 491

Qy        296 ATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 355
              |   || |||  ||    |    ||  | ||| || ||   || ||  |    |||  |
Db        490 ACTTTGTCACATTCATAAATGAAAAATCGAAATACACCTGGATAACACTTATTCAAACCA 431

Qy        356 GCGGTA 361
                | ||
Db        430 AAGATA 425


RESULT 9
AV058630
LOCUS       AV058630                296 bp    mRNA    linear   EST 23-JUN-1999
DEFINITION  AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA
            clone 1810054H02, mRNA sequence.
ACCESSION   AV058630
VERSION     AV058630.1  GI:5158377
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 296)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .296
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1810054H02"
                     /sex="male"
                     /tissue_type="pancreas"
                     /dev_stage="adult"
                     /clone_lib="Mus musculus pancreas C57BL/6J adult"
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 296;
  Best Local Similarity   53.4%;  Pred. No. 93;
  Matches   71;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy        251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCG 310
              |||  ||  | |||  | |  |  | ||  ||  |||  |  |||||  | ||  | |  |
Db         13 GTGGAGACCACAGTTTTTTCTAGGTCACCAAGGCTGGGTAAAGAAAAGAAGCATCAACTG 72

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              ||||||  | | | |||||     ||     | ||     || | | |  || | ||| |  |
Db         73 TCCAGTGAAGCACAAAAAATGAAGAAGAAAATGATTACCCAAAAGTTGGAATTAAGGTAG 132

Qy        371 CGCTGGTTAATTA 383
                | ||   || ||
Db        133 AGATGCCAAAATA 145
```

```
RESULT 10
AA746477/c
LOCUS       AA746477                 365 bp    mRNA    linear   EST 27-JAN-1998
DEFINITION  nw62e02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
            similar to TR:O00555 O00555 NEURONAL CALCIUM CHANNEL ALPHA 1A
            SUBUNIT ISOFORM A-1. ;contains element TAR1 repetitive element ;,
            mRNA sequence.
ACCESSION   AA746477
VERSION     AA746477.1  GI:2786463
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 661   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1251194"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 365;
  Best Local Similarity   56.9%;  Pred. No. 1e+02;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  ||||||   ||||||   | ||   | ||| |||  | ||| |||   |  ||||
Db        119 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGTGGCCAGGCCGGCCGGGCGCCACC 60

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |  ||  |  ||| ||||  ||   | ||   | | |   | | ||||  |
Db         59 AGCGGCCCTCGGAGGTACCCAGGCCCGCACGGCCGACGCTACTGGCCGG 11


RESULT 11
```

```
CD374421/c
LOCUS       CD374421                 558 bp    mRNA    linear   EST 19-AUG-2003
DEFINITION  TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)
            Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION   CD374421
VERSION     CD374421.1  GI:31229939
KEYWORDS    EST.
SOURCE      Wuchereria bancrofti
  ORGANISM  Wuchereria bancrofti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Wuchereria.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
  TITLE     Toward the characterization of the Wuchereria bancrofti
            microfilarial transcriptome with comparisons to those of Brugia
            malayi and Onchocerca volvulus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Thomas B. Nutman
            Laboratory of Parasitic Diseases
            NIH
            Building 4 Room 126, Bethesda, MD 20892-0425, USA
            Email: tnutman@niaid.nih.gov
            Seq primer: pBluescript T3.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Wuchereria bancrofti"
                     /mol_type="mRNA"
                     /db_xref="taxon:6293"
                     /clone="TNWbmfC2H3"
                     /dev_stage="microfilaria"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="Wuchereria bancrofti microfilaria cDNA
                     (SAW95SjL-WbMf)"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 85,000 microfilariae
                     isolated from the blood of an infected individual from
                     Guyana, South America and converted to double-stranded
                     cDNA using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNA pol I. The library has 2.9 x 10E5
                     independent recombinants and the average insert size is
                     ~1kb. The library was constructed by Sandra J. Laney. The
                     library is available from Dr.S.A.Williams, email:
                     genome@smith.edu."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 558;
  Best Local Similarity   60.7%;  Pred. No. 1.2e+02;
  Matches   65;  Conservative    0;  Mismatches   40;  Indels    2;  Gaps    1;

Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              ||| |||| |||||||| |      || ||||| |||| |    | ||  | |
Db        161 TGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCCCAAATTCTANGCT 102

Qy        136 AGCATTTATC--AGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||| |  ||  ||||  ||  | | |||| ||||||||| ||
Db        101 GACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55


RESULT 12
BU765156/c
LOCUS       BU765156                 559 bp    mRNA    linear   EST 10-OCT-2002
DEFINITION  sas23f11.y1 Gm-c1080 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1080-4869 5' similar to SW:D100_ARATH Q00874
            DNA-DAMAGE-REPAIR/TOLERATION PROTEIN DRT100 PRECURSOR ;, mRNA
            sequence.
ACCESSION   BU765156
VERSION     BU765156.1  GI:23733884
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
```

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 559)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: ccu@resgen.com web site: www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 448.
FEATURES Location/Qualifiers
source 1. .559
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1080-4869"
/tissue_type="Roots of 8 day old 'Bragg' supernodulating mutant NTS382 seedlings"
/dev_stage="8 days old"
/lab_host="DH10B"
/clone_lib="Gm-c1080"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The mRNA was isolated from roots of 8 day old 'Bragg' supernodulating mutant NTS382 seedlings that were infected with Bradyhizobium japonicum, strain USDA 110, 72 hours prior to harvest. Dr. Gary Stacey generously donated the tissue. The roots were flash-frozen in liquid nitrogen. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An 'anchor' nucleotide (V=A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first- strand synthesis primer was then restricted by digestion with XhoI from Promega (40U/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using Sephacryl S-500 High Resolution (Pharmacia Biotech) in a 2-mm diameter column and a bed volume of approximately 1ml. The column eluent was precipitated, redissolved, and ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). This library was constucted in the laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at Northern Arizona University."
ORIGIN

Query Match 7.4%; Score 33.8; DB 13; Length 559;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy        76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
             ||||| || | |    |||| | ||  |||||| ||| | |  || |||     ||||
Db       233 TGGGGTCGCAGGCGACGCCATACCAGCTCCGGCAACAGTACAGTGCCCGACCATGTGTTG 174

Qy       136 AGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGT 192
             |  ||     |||| |||||    |  || ||  |     | || | |||||||
Db       173 AAGATGCCCAAGTAGGGTTCGATCAGGGCAGCTTTGAAGGCCAAGAGTGCTGCCCGT 117


RESULT 13
AQ637195/c
LOCUS AQ637195 695 bp DNA linear GSS 17-JUN-1999
DEFINITION RPCI-11-481M9.TV RPCI-11 Homo sapiens genomic clone RPCI-11-481M9, genomic survey sequence.
ACCESSION AQ637195
VERSION AQ637195.1 GI:5099830
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 695)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-481M9.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .695
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7684616"
/db_xref="taxon:9606"
/clone="RPCI-11-481M9"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
ORIGIN

Query Match 7.4%; Score 33.8; DB 28; Length 695;
Best Local Similarity 49.2%; Pred. No. 1.4e+02;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```
Qy   275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
             | ||| | | || || | | || | |||| || | | || |
Db   291 TAACTAATACTTGTACACACAAAAGAGGATTAATGATCAAGTGTCACATTCTAGACACGT 232

Qy   335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGG 394
             || |||| || | | | || | || |||| | | | ||| |
Db   231 ATTTTACAGTTGTAACTCAAGTGCATACAGATGCACAGCTGTACACTCACACTCATACAG 172

Qy   395 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAGTATT 454
             | | ||| | || ||| ||| | ||||| || || | ||| || || | |
Db   171 CCATACATGTGGACACACATTAAAAGTCACATATAAGCATGGACACAAGTATACATTCAT 112
```
```

```
Qy          455 A 455
                |
Db          111 A 111


RESULT 14
CG178280
LOCUS       CG178280                941 bp    DNA     linear   GSS 21-AUG-2003
DEFINITION  PUJDF79TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0641N14,
            genomic survey sequence.
ACCESSION   CG178280
VERSION     CG178280.1  GI:34069341
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 941)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUJDF79TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .941
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa0641N14"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 29;  Length 941;
  Best Local Similarity   51.0%;  Pred. No. 1.6e+02;
  Matches   80;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | ||||  | | || || | || || | || ||  | |||  ||     | | ||||| |
Db        236 TTGCAGTCTACACATCAGTGGTCGTATATGCCACAGTTCCTGCACCGGCCTTTCCGCCGC 295

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGA 136
              ||  |   | | ||  |||  ||| | || |  |     |  ||| ||||| ||   ||
Db        296 GGAGTACCCTCCGTCGTCACCACGACCGCCGGTCTTCGACTTGCCGGGACTGTACCCCGA 355

Qy        137 GCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGC 173
                 ||  ||   || | ||  |     | | ||||
Db        356 CTTCTTCTCGCCACCGCTCAACGGCGACGACTCTGGC 392


RESULT 15
CNS02GQP/c
LOCUS       CNS02GQP                1098 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            138K23 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL196666
VERSION     AL196666.1  GI:7834816
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1098)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1098
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="138K23"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG138AF12SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 29;  Length 1098;
  Best Local Similarity   55.8%;  Pred. No. 1.7e+02;
  Matches   53;  Conservative    5;  Mismatches   37;  Indels    0;  Gaps    0;

Qy         19 GCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGG 78
              | ||   | || | |   | | || ||||    | |||||||   | |||| |:|
Db        773 GAAGTCCTGGCTGAACAGGGGGATTTTGGCGTAACGCTGGCTGAAATGGTTCAAMACRAT 714

Qy         79 GGCGGCGGCGGTAATCATAACGGCGGCGGCAATAG 113
              :  ||||||| |:||| |:||| |: || | || |
Db        713 RAAGGCGGCGTTMATCCTMACGCCRACGCCGATGG 679


Search completed: March 18, 2004, 20:11:25
Job time : 2034.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		404.8	88.8	456	6	I44909	I44909 Sequence 58
2		404.8	88.8	2067	1	SEU43280	U43280 Salmonella
3		396.8	87.0	5103	1	STAJ2301	AJ002301 Salmonell
4		396.8	87.0	22411	1	AE008749	AE008749 Salmonell
5		395.2	86.7	254050	1	AL627269	AL627269 Salmonell
6	c	395.2	86.7	301983	1	AE016840	AE016840 Salmonell
7		389.4	85.4	1048	1	STAGFBA	AJ000514 Salmonell
8		287.2	63.0	361	6	I44908	I44908 Sequence 56
9		281.6	61.8	2889	1	CSP515700	AJ515700 Citrobact
10		229.2	50.3	2920	1	CFR515701	AJ515701 Citrobact
11		229	50.2	230	1	SEU53207	U53207 Salmonella
12		227.2	49.8	4680	1	ECCSGABDG	X90754 E.coli csgG
13		227.2	49.8	10346	1	AE000205	AE000205 Escherich
14		227.2	49.8	15047	1	D90741	D90741 Escherichia
15		225.6	49.5	456	6	AX814811	AX814811 Sequence
16		225.6	49.5	648	1	ECOCSGAA	L04979 Escherichia
17		220.6	48.4	306358	1	AE016759	AE016759 Escherich
18		217.4	47.7	1711	1	AF275733	AF275733 Escherich
19		217.4	47.7	10190	1	AE005315	AE005315 Escherich
20		217.4	47.7	327773	1	AP002554	AP002554 Escherich
21		188.8	41.4	2883	1	ESA515702	AJ515702 Enterobac
22		163.2	35.8	10370	1	AE015131	AE015131 Shigella
23		163.2	35.8	292504	1	AE016981	AE016981 Shigella
24	c	140.2	30.7	437	1	AF237726	AF237726 Shigella
25		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
26		64.4	14.1	19201	1	D90742	D90742 Escherichia
27		44.8	9.8	72	6	AX814798	AX814798 Sequence
28	c	44.8	9.8	2000	6	AX655393	AX655393 Sequence
29		39.4	8.6	956	8	AK058943	AK058943 Oryza sat
30		39.4	8.6	2720	8	AK102629	AK102629 Oryza sat
31		39.4	8.6	152448	8	AP004339	AP004339 Oryza sat
32		38.8	8.5	13264	1	AE001012	AE001012 Archaeogl
33		38.2	8.4	125020	9	AF429315	AF429315 Homo sapi
34	c	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
35	c	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
36	c	36	7.9	168267	2	AC146009	AC146009 Pan trogl
37		35.8	7.9	172600	9	AL359644	AL359644 Human DNA
38	c	35.8	7.9	193168	2	AC011969	AC011969 Homo sapi
39		35.6	7.8	301214	1	AE016786	AE016786 Pseudomon
40	c	35	7.7	117541	2	AC104285	AC104285 Oryza sat
41		34.4	7.5	560	8	AY247273	AY247273 Medicago
42		34.4	7.5	1553	1	HAARR16S	M59123 Haloanaerob
43		34.4	7.5	2751	6	AX416572	AX416572 Sequence
44		34.4	7.5	151828	2	AC146008	AC146008 Pan trogl
45	c	34.4	7.5	160449	10	AC130549	AC130549 Mus muscu

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity 93.0%;  Pred. No. 8.9e-106;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||        |  || | ||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match           88.8%;  Score 404.8;  DB 1;  Length 2067;
  Best Local Similarity 93.0%;  Pred. No. 1e-105;
```

```
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||         |  || | ||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301                5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match           87.0%;  Score 396.8;  DB 1;  Length 5103;
  Best Local Similarity  91.9%;  Pred. No. 2.3e-103;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
     RBS                complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene               complement(2580. .4088)
                        /gene="STM1128"
     CDS                complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS                complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene               complement(4413. .5099)
                        /gene="STM1129"
     CDS                complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
     RBS                complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene               5599. .6759
                        /gene="STM1130"
     CDS                5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene               6793. .7497
                        /gene="STM1131"
     RBS                6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           87.0%;  Score 396.8;  DB 1;  Length 22411;
  Best Local Similarity   91.9%;  Pred. No. 2.6e-103;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||| |
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||          |  || | ||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             86.7%;  Score 395.2;  DB 1;  Length 254050;
  Best Local Similarity   91.7%;  Pred. No. 9.6e-103;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||         |  || | ||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361
```

RESULT 6
AE016840/c
LOCUS AE016840 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of 16 of the complete genome.
ACCESSION AE016840 AE014613
VERSION AE016840.1 GI:29137797
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi Ty2
ORGANISM Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 301983)
AUTHORS Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and CT18
JOURNAL J. Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE 22531367
PUBMED 12644504
REFERENCE 2 (bases 1 to 301983)
AUTHORS Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES Location/Qualifiers

```
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.7%;  Score 395.2;  DB 1;  Length 301983;
Best Local Similarity   91.7%;  Pred. No. 9.8e-103;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||| |
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                 ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             85.4%;  Score 389.4;  DB 1;  Length 1048;
  Best Local Similarity   91.0%;  Pred. No. 2.7e-101;
  Matches  414;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||||||||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||||  |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             63.0%;  Score 287.2;  DB 6;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 8.4e-72;
  Matches  289;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
```

```
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           61.8%;  Score 281.6;  DB 1;  Length 2889;
  Best Local Similarity  77.6%;  Pred. No. 4.3e-70;
  Matches  354;  Conservative    0;  Mismatches   99;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||| |||||||||||||| |||||||||||| ||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||    |      ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||    |  |||||| |||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||    ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||| ||||||||||  ||||| || || ||||||||  |||| |||| |
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              |      ||     | || | |||       | |  || | |||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571


RESULT 10
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN

  Query Match           50.3%;  Score 229.2;  DB 1;  Length 2920;
  Best Local Similarity  70.6%;  Pred. No. 5.9e-55;
  Matches  322;  Conservative    0;  Mismatches  128;  Indels    6;  Gaps    1;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||| |  |||| || || || || || ||     ||| ||||  ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || |||||| |||||||| ||     ||||| || ||  |||  ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||| ||| | ||||| |||  ||||  |||| || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||| ||||| |||||||| || |||||  | |||||||  ||   ||| |  | ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              |          | ||||  ||   |  | | | | ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || |||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

RESULT 11
SEU53207
LOCUS SEU53207 230 bp DNA linear BCT 01-MAY-1996
DEFINITION Salmonella enteritidis SEF17 fimbrin (agfA) gene, partial cds.
ACCESSION U53207
VERSION U53207.1 GI:1293677
KEYWORDS .
SOURCE Salmonella enteritidis
ORGANISM Salmonella enteritidis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 230)
AUTHORS Cox,J.M., Eglezos,S. and Woolcock,J.B.
TITLE Virulence of Salmonella enteritidis in chickens correlates with colony morphology and expression of SEF17 fimbriae
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 230)
AUTHORS Cox,J.M., Eglezos,S. and Woolcock,J.B.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1996) J.M. Cox, Food Science and Technology, The University of New South Wales, Sydney, NSW 2052, Australia
FEATURES Location/Qualifiers
source 1. .230
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="SE30"
/db_xref="taxon:592"
/note="amplification product"
gene 1. .230
/gene="agfA"
CDS <1. .>230
/gene="agfA"
/codon_start=1
/transl_table=11
/product="SEF17 fimbrin"
/protein_id="AAA98671.1"
/db_xref="GI:1293678"
/translation="GNHXGGGNSSGPDSTLSIYQYGSANAALALQSDARKSETTITQS
GYGNGADVGQGADNSTIELTQNGFRNNATIDQ"

ORIGIN

Query Match 50.2%; Score 229; DB 1; Length 230;
Best Local Similarity 99.6%; Pred. No. 5.3e-55;
Matches 229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         88 GGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 147
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTAATCATAANGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 60

Qy        148 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACC 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACC 120

Qy        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 180

Qy        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 230
```

RESULT 12
ECCSGABDG
LOCUS ECCSGABDG 4680 bp DNA linear BCT 07-JUL-2002
DEFINITION E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION X90754
VERSION X90754.1 GI:1147558
KEYWORDS csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene; orfC gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
TITLE Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12
JOURNAL Mol. Microbiol. 18 (4), 661-670 (1995)
MEDLINE 96414468
PUBMED 8817489
REFERENCE 2 (bases 1 to 4680)
AUTHORS Hammar,M.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet, Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm, SWEDEN
FEATURES Location/Qualifiers
source 1. .4680
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/map="23.15 minutes"
source 1. .3140
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="MC4100"
/db_xref="taxon:562"
source 3141. .4680
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="W3110"
/db_xref="taxon:562"
repeat_region <1. .>4680
/insertion_seq="IS2 (partial)"
gene complement(133. .966)
/gene="csgG"
CDS complement(133. .966)
/gene="csgG"
/codon_start=1

```
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
gene                 complement(993. .1409)
                     /gene="csgF"
CDS                  complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene                 complement(1434. .1823)
                     /gene="csgE"
CDS                  complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
gene                 complement(1828. .2478)
                     /gene="csgD"
CDS                  complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
misc_feature         complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
misc_feature         3140
                     /note="transcriptional start site for csgBA operon"
gene                 3233. .3688
                     /gene="csgB"
CDS                  3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
gene                 3729. .4184
                     /gene="csgA"
CDS                  3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
sig_peptide          3729. .3788
                     /gene="csgA"
gene                 4243. .4575
                     /gene="orfC"
CDS                  4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match             49.8%;  Score 227.2;  DB 1;  Length 4680;
  Best Local Similarity   68.6%;  Pred. No. 2.3e-54;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || |||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||    ||| |||||||||    ||     ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||    ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || |||||||||  ||||| || || || || ||    || |  | |
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||  |||  |||         |  |  |   ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 13
AE000205
```

```
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
```

```
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

  Query Match           49.8%;  Score 227.2;  DB 1;  Length 10346;
  Best Local Similarity   68.6%;  Pred. No. 2.5e-54;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || |||||  || |  |||||||||||||| ||    || |||||  |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||   | ||    | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||

Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || |||||||||| ||||| || || || || ||    || |  | |
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                    ||  |||  |||        |  |  |   ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448


RESULT 14
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
```

```
            Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
          Headed by:
            Name: Takashi Horiuchi
            Address: National Institute of Basic Biology, Okazaki, 444, Japan
            E-mail: kishori@nibb.ac.jp
          Information operator:
            Name: Hirotada Mori
            Address: NARA Institute of Science and Technology,
                     Ikoma, 630-01, Japan
            E-mail: hmori@gtc.aist-nara.ac.jp
          URL:
            The Japan E. coli genome database
            http:bsw3.aist-nara.ac.jp.
FEATURES            Location/Qualifiers
     source         1. .15047
                    /organism="Escherichia coli K12"
                    /mol_type="genomic DNA"
                    /strain="K12"
                    /db_xref="taxon:83333"
                    /map="23.7-24.0 min"
                    /clone="Kohara clone #231"
                    /note="Nucleotide position 1097101-1112147 from the
                    initiation site of ThrA (0 min.).~This clone is from
                    Kohara lambda miniset library."
     CDS            <1. .276
                    /note="unnamed protein product; ORF_ID:o230#5"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="BAA35822.1"
                    /db_xref="GI:4062605"
                    /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
                    WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     CDS            1152. .2090
                    /note="ORF_ID:o231#1
                    similar to SwissProt Accession Number P45637"
                    /codon_start=1
                    /transl_table=11
                    /product="Hypothetical 33.0 kd protein in proB-proA
                    intergenic region."
                    /protein_id="BAA35823.1"
                    /db_xref="GI:4062606"
                    /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
                    WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
                    AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
                    PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
                    GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
                    AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     CDS            2145. .2882
                    /note="unnamed protein product; ORF_ID:o231#2"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="BAA35824.1"
                    /db_xref="GI:4062607"
                    /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                    EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                    PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                    SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
                    ILNVSPRRLLNFLESRGMAPIAEFADL"
     CDS            2906. .3460
                    /note="ORF_ID:o231#3
                    similar to PIR Accession Number F64035"
                    /codon_start=1
                    /transl_table=11
                    /product="Hypothetical protein HI1543"
                    /protein_id="BAA35825.1"
                    /db_xref="GI:4062608"
                    /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                    DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                    ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                    RTMAPLTRDAISAMWDELEEDSEE"
     CDS            3514. .4053
                    /note="unnamed protein product; ORF_ID:o231#4"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="BAA35826.1"
                    /db_xref="GI:4062609"
                    /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                    SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                    IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                    LAARSAKTAHREQEIKNKA"
     gene           complement(4117. .4950)
                    /gene="csgG"
     CDS            complement(4117. .4950)
                    /gene="csgG"
                    /note="ORF_ID:o231#5
                    similar to SwissProt Accession Number P52103"
                    /codon_start=1
                    /transl_table=11
                    /product="Curli production assembly/transport component
                    CsgG precursor."
                    /protein_id="BAA35827.1"
                    /db_xref="GI:4062610"
                    /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                    PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                    NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                    IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                    VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                    S"
     gene           complement(4977. .5393)
                    /gene="csgF"
     CDS            complement(4977. .5393)
                    /gene="csgF"
                    /note="ORF_ID:o231#6
                    similar to SwissProt Accession Number P52104"
                    /codon_start=1
                    /transl_table=11
                    /product="Curli production assembly/transport component
                    CsgF precursor."
                    /protein_id="BAA35828.1"
                    /db_xref="GI:4062611"
                    /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                    AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                    DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene           complement(5418. .5807)
                    /gene="csgE"
     CDS            complement(5418. .5807)
                    /gene="csgE"
                    /note="ORF_ID:o231#7
                    similar to SwissProt Accession Number P52105"
                    /codon_start=1
                    /transl_table=11
                    /product="Curli production assembly/transport component
                    CsgE precursor."
                    /protein_id="BAA35829.1"
                    /db_xref="GI:4062612"
                    /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                    RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                    IQTEEALNRRQINQALLSTGDLAHDEF"
     gene           complement(5812. .6462)
                    /gene="csgD"
     CDS            complement(5812. .6462)
                    /gene="csgD"
                    /note="ORF_ID:o231#8
                    similar to SwissProt Accession Number P52106"
                    /codon_start=1
                    /transl_table=11
                    /product="Probable csgAB operon transcriptional regulatory
                    protein."
                    /protein_id="BAA35830.1"
                    /db_xref="GI:4062613"
                    /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                    RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                    HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                    KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene           7217. .7672
```

```
                 /gene="csgB"
     CDS         7217. .7672

  Query Match           49.8%;  Score 227.2;  DB 1;  Length 15047;
  Best Local Similarity   68.6%;  Pred. No. 2.6e-54;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || |||||||||  ||||| || || || ||  ||| |  || | | |
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                 ||  |||  |||       |  |  |   ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 15
AX814811
LOCUS       AX814811                456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           49.5%;  Score 225.6;  DB 6;  Length 456;
  Best Local Similarity   68.4%;  Pred. No. 5.5e-54;
  Matches  312;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || |||||||||  ||||| || || || ||  ||| |  || | | |
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                 ||  |||  |||       |  |  |   ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


Search completed: March 18, 2004, 15:02:45
Job time : 2312.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64623	Aac64623 AgfA::PT3
	2	404.8	88.8	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	404.8	88.8	456	2	AAT74142	Aat74142 Salmonell
	4	404.8	88.8	456	3	AAC64617	Aac64617 Salmonell
	5	382.4	83.9	456	3	AAC64628	Aac64628 AgfA::PT3
	6	376	82.5	456	3	AAC64622	Aac64622 AgfA::PT3
	7	361.6	79.3	456	3	AAC64626	Aac64626 AgfA::PT3
	8	355.2	77.9	456	3	AAC64625	Aac64625 AgfA::PT3
	9	353.6	77.5	456	3	AAC64629	Aac64629 AgfA::PT3
	10	352	77.2	456	3	AAC64624	Aac64624 AgfA::PT3
	11	350.4	76.8	456	3	AAC64630	Aac64630 AgfA::PT3
	12	350.4	76.8	456	3	AAC64627	Aac64627 AgfA::PT3
	13	348.8	76.5	456	3	AAC64631	Aac64631 AgfA::PT3
	14	287.2	63.0	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	287.2	63.0	361	2	AAT74141	Aat74141 Salmonell
	16	225.6	49.5	456	3	AAC64619	Aac64619 Escherich
	17	225.6	49.5	456	9	ACF36153	Acf36153 E. coli C
	18	218	47.8	646	2	AAQ62647	Aaq62647 Fibronect
	19	156	34.2	369	2	AAQ62646	Aaq62646 FNB curli
	20	49	10.7	78	3	AAC64610	Aac64610 AgfA (SEF
c	21	48.4	10.6	78	3	AAC64609	Aac64609 AgfA (SEF
	22	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	23	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
c	24	44.8	9.8	2000	7	ADA71938	Ada71938 Rice gene
	25	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
c	26	41.6	9.1	78	3	AAC64605	Aac64605 SefA (SEF
	27	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
	28	40.6	8.9	78	3	AAC64606	Aac64606 SefA (SEF
	29	37.6	8.2	456	3	AAC64618	Aac64618 Salmonell
	30	37.4	8.2	456	3	AAC64620	Aac64620 Escherich
	31	35.4	7.8	100	7	ACD68807	Acd68807 E. coli K
c	32	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	33	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
	34	35	7.7	100	7	ACD68808	Acd68808 E. coli K
	35	34.8	7.6	1083	5	AAS76745	Aas76745 DNA encod
	36	34.4	7.5	2751	6	ABQ70750	Abq70750 Listeria
	37	34.2	7.5	2000	7	ADA71938	Ada71938 Rice gene
c	38	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	39	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	40	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	41	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	42	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	43	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c	44	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a
	45	33.4	7.3	78	9	ACF36151	Acf36151 E. coli C

ALIGNMENTS

RESULT 1
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.2e-138;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 2; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.2e-121;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||        |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 2;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.2e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||      |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

```
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 3;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.2e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||       |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             83.9%;  Score 382.4;  DB 3;  Length 456;
  Best Local Similarity   89.9%;  Pred. No. 2.4e-114;
  Matches  410;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||| |||       ||   ||    || |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              |  |  |       ||  |  ||         |  || | ||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 6
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match            82.5%;  Score 376;  DB 3;  Length 456;
  Best Local Similarity  89.0%;  Pred. No. 2.9e-112;
  Matches  406;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||  | ||||  ||||  |  | |  | ||  |  | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match             79.3%;  Score 361.6;  DB 3;  Length 456;
  Best Local Similarity   87.1%;  Pred. No. 1.5e-107;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| ||| |  |||||| | |  | | ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||   | ||||  ||          |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 1.8e-105;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |   ||  |     | | ||  | |    || | ||| ||||||
```

```
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||        |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             77.5%;  Score 353.6;  DB 3;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 6.1e-105;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||        |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64624
ID   AAC64624 standard; DNA; 456 BP.
XX
AC   AAC64624;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 2e-104;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | |   | |   |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||       |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 6.8e-104;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | |||  |    | | ||  |   |     ||| ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||   | ||||  ||      |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 6.8e-104;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||  |  |       | || |    || | |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||      |||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||   | ||||  ||        |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match           76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 2.3e-103;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||  || |||  |   |           |   | |  ||   |
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||          |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             63.0%;  Score 287.2;  DB 2;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 2.7e-83;
  Matches  289;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag= b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag= c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag= d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag= e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             63.0%;  Score 287.2;  DB 2;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 2.7e-83;
  Matches  289;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292


Search completed: March 18, 2004, 09:11:56
Job time : 289.405 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	404.8	88.8	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	287.2	63.0	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	33	7.2	7766	4	US-09-125-619-3	Sequence 3, Appli
c	6	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	7	31.2	6.8	1767	1	US-08-363-475-18	Sequence 18, Appl
c	8	31.2	6.8	2284	1	US-08-363-475-21	Sequence 21, Appl
	9	31	6.8	1750	1	US-08-415-823-1	Sequence 1, Appli
	10	31	6.8	1750	2	US-09-086-662-1	Sequence 1, Appli
c	11	30.6	6.7	1147	1	US-08-417-103-15	Sequence 15, Appl
c	12	30.6	6.7	1351	1	US-07-816-283-5	Sequence 5, Appli
c	13	30.6	6.7	1351	1	US-08-417-103-5	Sequence 5, Appli
c	14	30.6	6.7	1351	4	US-09-016-434-1303	Sequence 1303, Ap
c	15	30.6	6.7	1797	4	US-09-134-000C-1552	Sequence 1552, Ap
c	16	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	17	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	18	30.4	6.7	7183	4	US-09-081-149-9	Sequence 9, Appli
c	19	30.4	6.7	7183	4	US-09-081-149-10	Sequence 10, Appl
c	20	30.2	6.6	2007	4	US-09-643-597-153	Sequence 153, App
c	21	30.2	6.6	2007	4	US-09-480-884A-153	Sequence 153, App
c	22	30.2	6.6	2007	4	US-09-542-615A-153	Sequence 153, App
c	23	30.2	6.6	2007	4	US-09-606-421B-153	Sequence 153, App
c	24	30.2	6.6	2007	4	US-09-221-107-153	Sequence 153, App
c	25	30.2	6.6	2148	4	US-09-643-597-154	Sequence 154, App
c	26	30.2	6.6	2148	4	US-09-480-884A-154	Sequence 154, App
c	27	30.2	6.6	2148	4	US-09-542-615A-154	Sequence 154, App
c	28	30.2	6.6	2148	4	US-09-606-421B-154	Sequence 154, App
c	29	30.2	6.6	2148	4	US-09-221-107-154	Sequence 154, App
	30	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	31	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	32	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
c	33	29.8	6.5	3612	4	US-09-068-506-2	Sequence 2, Appli
c	34	29.8	6.5	3690	4	US-09-016-434-1234	Sequence 1234, Ap
c	35	29.8	6.5	4035	4	US-09-016-434-1369	Sequence 1369, Ap
c	36	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
c	37	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	38	29.4	6.4	759	2	US-08-853-659A-18	Sequence 18, Appl
c	39	29.4	6.4	918	2	US-08-853-659A-22	Sequence 22, Appl
c	40	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	41	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	42	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	43	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	44	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
c	45	29.4	6.4	9208	4	US-09-068-506-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.5e-131;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||        |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             63.0%;  Score 287.2;  DB 1;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 1.8e-90;
  Matches  289;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292
```

```
RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
;  GENERAL INFORMATION:
;    APPLICANT:  Kim, Jihyun Francis
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;    TITLE OF INVENTION:  ERWINIA AMYLOVORA AND ITS USE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,927
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,108
;      FILING DATE:  06-AUG-1977
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1581
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1344 base pairs
```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-120-927-1

Query Match 7.5%; Score 34; DB 3; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.13;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/41 (EBC-003)
; CURRENT APPLICATION NUMBER: US/09/431,614
; CURRENT FILING DATE: 1999-11-02
; EARLIER APPLICATION NUMBER: 60/107,243
; EARLIER FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-431-614-5

Query Match 7.5%; Score 34; DB 4; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.13;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234
; CURRENT APPLICATION NUMBER: US/09/125,619
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match 7.2%; Score 33; DB 4; Length 7766;
Best Local Similarity 47.8%; Pred. No. 0.77;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||     ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |   |  | ||||||||||  | |||   |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |  || ||| |   |    | ||| |          | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATATGAT 357
              || ||||        |||||
Db       4612 TTTGCTGTGAAGAGTAATGAT 4632
```

RESULT 6
5486473-3/c
;Patent No. 5486473
; APPLICANT: FUKITA, HIROYUKI;YOSHIDA, IWAO;TAKAGI, MITSUO;
;MANABE, SADAO;FUKAI, KONOSUKE
; TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,049
; FILING DATE: 09-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 809,255
; FILING DATE: 18-DEC-1991
; APPLICATION NUMBER: 279,685
; FILING DATE: 05-DEC-1988
; APPLICATION NUMBER: 932,419
; FILING DATE: 19-NOV-1986
;SEQ ID NO:3:
; LENGTH: 1491
5486473-3

Query Match 6.8%; Score 31.2; DB 6; Length 1491;
Best Local Similarity 50.0%; Pred. No. 1.3;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||  |  || ||| | |  | |||    | | | || | | |  | | |  | ||  |
Db        204 CAAGTCACTGACCGAAGCTAGGTAACAGTAACTGCGCACATCTGCGAGGTTGGCTGCTTC 145

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||   ||  ||   ||| ||   |||   |   ||  |||||  || || |
Db        144 CATGTTCATCATTTTGACATCAATGGTTGGCTTGTCTTTTGACATTATGGTCACACAACT 85

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
              |  |||  |   |||||  |||  |   | ||||
Db         84 ATCGCCTTCCAGTACCAGATCAACCCATGTAGCTCC 49
```

```
RESULT 7
US-08-363-475-18/c
; Sequence 18, Application US/08363475
; Patent No. 5516679
;  GENERAL INFORMATION:
;    APPLICANT:  Chiang, Shu-Jen
;    APPLICANT:  Burnett Jr., Willian V.
;    APPLICANT:  Tonzi, Sean M.
;    TITLE OF INVENTION:  PENICILLIN V AMIDOHYDROLASE GENE FROM
;    TITLE OF INVENTION:  FUSARIUM OXYSPORUM
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas R. Savitsky
;      STREET:  P.O. Box 4000
;      CITY:  Princeton
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/363,475
;      FILING DATE:  23-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savitsky, Thomas R.
;      REGISTRATION NUMBER:  31,661
;      REFERENCE/DOCKET NUMBER:  ON-0134
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 252 4956
;      TELEFAX:  (609) 252 4526
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1767 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1521
US-08-363-475-18

  Query Match             6.8%;  Score 31.2;  DB 1;  Length 1767;
  Best Local Similarity   55.6%;  Pred. No. 1.4;
  Matches   60;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        278 CTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATA 337
              | |  ||| ||||| |    || |||  | |||||| |   ||  ||   |      | |
Db       1391 CCCCCAATCGTTTCTGGCGCAACGCCTTCCTCGACCCATTTCACAATAGCCAGAAGCACA 1332

Qy        338 TTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATG 385
              ||| | || |||    ||     |  ||| || | |||||| || |||
Db       1331 TTAGTATCTGCCGTCTTGCCGCCCATGTTTCCGGCGTTGTTTCCAATG 1284


RESULT 8
US-08-363-475-21/c
; Sequence 21, Application US/08363475
; Patent No. 5516679
;  GENERAL INFORMATION:
;    APPLICANT:  Chiang, Shu-Jen
;    APPLICANT:  Burnett Jr., Willian V.
;    APPLICANT:  Tonzi, Sean M.
;    TITLE OF INVENTION:  PENICILLIN V AMIDOHYDROLASE GENE FROM
;    TITLE OF INVENTION:  FUSARIUM OXYSPORUM
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas R. Savitsky
;      STREET:  P.O. Box 4000
;      CITY:  Princeton
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/363,475
;      FILING DATE:  23-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savitsky, Thomas R.
;      REGISTRATION NUMBER:  31,661
;      REFERENCE/DOCKET NUMBER:  ON-0134
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 252 4956
;      TELEFAX:  (609) 252 4526
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2284 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  241..1836
US-08-363-475-21

  Query Match             6.8%;  Score 31.2;  DB 1;  Length 2284;
  Best Local Similarity   55.6%;  Pred. No. 1.6;
  Matches   60;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        278 CTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATA 337
              | |  ||| ||||| |    || |||  | |||||| |   ||  ||   |      | |
Db       1706 CCCCCAATCGTTTCTGGCGCAACGCCTTCCTCGACCCATTTCACAATAGCCAGAAGCACA 1647

Qy        338 TTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATG 385
              ||| | || |||    ||     |  ||| || | |||||| || |||
Db       1646 TTAGTATCTGCCGTCTTGCCGCCCATGTTTCCGGCGTTGTTTCCAATG 1599


RESULT 9
US-08-415-823-1
; Sequence 1, Application US/08415823
; Patent No. 5759538
;  GENERAL INFORMATION:
;    APPLICANT:  Donovan, William P.
;    APPLICANT:  Tan, Yuping
;    TITLE OF INVENTION:  BACILLUS THURINGIENSIS apr AND npr
;    TITLE OF INVENTION:  PROTEASE GENES AND ALKALINE PROTEASE DEFICIENT AND NEUTRAL
;    TITLE OF INVENTION:  PROTEASE DEFICIENT B.T. STRAINS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Panitch Schwarze Jacobs & Nadel, P.C.  c/o
;      ADDRESSEE:  A.S. Nadel
;      STREET:  1601 Market Street, 36th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103-2398
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,823
```

```
;       FILING DATE:
;       CLASSIFICATION:  435
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1750 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
US-08-415-823-1

  Query Match             6.8%;  Score 31;  DB 1;  Length 1750;
  Best Local Similarity   48.1%;  Pred. No. 1.6;
  Matches   88;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        254 CGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACC 313
              || | |||  |   |  | ||||    ||||    |  |||  ||     |  ||| |
Db        449 CGTACAATTTTCAAAAGGTACTGTAAAAGAAAAAATAAAGAGCTATAAAAATAATCCAGA 508

Qy        314 AGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTG 373
               ||||||     | |||     ||||||  ||   |  | ||  |  |     |||| |
Db        509 TGTGGAATATGCAGAACCGAATTATTACGTTCACGCCTTTTGGACTCCAAACGACCCATA 568

Qy        374 TTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACA 433
              || | |      | | ||   |   |||  | | | | |   || ||||  | | ||||
Db        569 TTTTAATAATCAATACGGGTTACAAAAGATTCAAGCTCCACAAGCTTGGGATAGCCAACG 628

Qy        434 ACG 436
              | ||
Db        629 AAG 631


RESULT 10
US-09-086-662-1
; Sequence 1, Application US/09086662
; Patent No. 5962264
;  GENERAL INFORMATION:
;    APPLICANT:  Donovan, William P.
;    APPLICANT:  Tan, Yuping
;    TITLE OF INVENTION:  BACILLUS THURINGIENSIS apr AND npr
;    TITLE OF INVENTION:  PROTEASE GENES AND ALKALINE PROTEASE DEFICIENT AND NEUTRAL
;    TITLE OF INVENTION:  PROTEASE DEFICIENT B.T. STRAINS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Panitch Schwarze Jacobs & Nadel, P.C.  c/o
;      ADDRESSEE:  A.S. Nadel
;      STREET:  1601 Market Street, 36th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103-2398
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/086,662
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/415,823
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1750 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
US-09-086-662-1

  Query Match             6.8%;  Score 31;  DB 2;  Length 1750;
  Best Local Similarity   48.1%;  Pred. No. 1.6;
  Matches   88;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        254 CGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACC 313
              || | |||  |   |  | ||||    ||||    |  |||  ||     |  ||| |
Db        449 CGTACAATTTTCAAAAGGTACTGTAAAAGAAAAAATAAAGAGCTATAAAAATAATCCAGA 508

Qy        314 AGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTG 373
               ||||||     | |||     ||||||  ||   |  | ||  |  |     |||| |
Db        509 TGTGGAATATGCAGAACCGAATTATTACGTTCACGCCTTTTGGACTCCAAACGACCCATA 568

Qy        374 TTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACA 433
              || | |      | | ||   |   |||  | | | | |   || ||||  | | ||||
Db        569 TTTTAATAATCAATACGGGTTACAAAAGATTCAAGCTCCACAAGCTTGGGATAGCCAACG 628

Qy        434 ACG 436
              | ||
Db        629 AAG 631


RESULT 11
US-08-417-103-15/c
; Sequence 15, Application US/08417103
; Patent No. 5723299
;  GENERAL INFORMATION:
;    APPLICANT:  Bell, Graeme I.
;    APPLICANT:  Yamada, Yuichiro
;    APPLICANT:  Seino, Susumu
;    TITLE OF INVENTION:  SOMATOSTATIN RECEPTORS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/417,103
;      FILING DATE:  05-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/816,283
;      FILING DATE:  01-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mark B.
;      REGISTRATION NUMBER:  37,259
;      REFERENCE/DOCKET NUMBER:  ARCD:144
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1147 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  15..1121
US-08-417-103-15

  Query Match             6.7%;  Score 30.6;  DB 1;  Length 1147;
  Best Local Similarity   48.1%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
              |||| |   |||  |  | |  |  |||||||  || ||||  | |  |  |||
```

```
Db        302 CAGACCCAGCATGAAGAGCTCATCTGCGATGGCCAGGTTGAGGATGTAAATGTTGGTGAT 243

Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333

Db        242 GGTCTTCATCTTGGCATAGCGGAGGATGACATAAATGACAAGTGTGTTGCCACACAACCC 183

Qy        334 GATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393

Db        182 AATGATGCAGACCACAAAATAGATGAATGTGAGGACTGCATTGCTTGTCAGGTCATAGTA 123

Qy        394 C 394

Db        122 C 122
```

RESULT 12
US-07-816-283-5/c
; Sequence 5, Application US/07816283
; Patent No. 5436155
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamada, Yuichiro
; APPLICANT: Seino, Susumu
; TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/816,283
; FILING DATE: 19911231
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDaniel, C. Steven
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1351 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-07-816-283-5

Query Match 6.7%; Score 30.6; DB 1; Length 1351;
Best Local Similarity 48.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273

Db        370 CAGACCCAGCATGAAGAGCTCATCTGCGATGGCCAGGTTGAGGATGTAAATGTTGGTGAT 311

Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333

Db        310 GGTCTTCATCTTGGCATAGCGGAGGATGACATAAATGACAAGTGTGTTGCCACACAACCC 251

Qy        334 GATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393

Db        250 AATGATGCAGACCACAAAATAGATGAATGTGAGGACTGCATTGCTTGTCAGGTCATAGTA 191

Qy        394 C 394

Db        190 C 190
```

RESULT 13
US-08-417-103-5/c
; Sequence 5, Application US/08417103
; Patent No. 5723299
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamada, Yuichiro
; APPLICANT: Seino, Susumu
; TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,103
; FILING DATE: 05-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/816,283
; FILING DATE: 01-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 83..1189
US-08-417-103-5

Query Match 6.7%; Score 30.6; DB 1; Length 1351;
Best Local Similarity 48.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273

Db        370 CAGACCCAGCATGAAGAGCTCATCTGCGATGGCCAGGTTGAGGATGTAAATGTTGGTGAT 311

Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333

Db        310 GGTCTTCATCTTGGCATAGCGGAGGATGACATAAATGACAAGTGTGTTGCCACACAACCC 251

Qy        334 GATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393

Db        250 AATGATGCAGACCACAAAATAGATGAATGTGAGGACTGCATTGCTTGTCAGGTCATAGTA 191

Qy        394 C 394

Db        190 C 190
```

RESULT 14

```
US-09-016-434-1303/c
; Sequence 1303, Application US/09016434
; Patent No. 6500938
;  GENERAL INFORMATION:
;    APPLICANT:  Janice Au-Young
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF SIGNALING
;    TITLE OF INVENTION:  PATHWAY GENE EXPRESSION
;    NUMBER OF SEQUENCES:  1490
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/016,434
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0002 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  1303:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1351 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GENBANK
;      CLONE:  g307435
US-09-016-434-1303

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 1351;
  Best Local Similarity   48.1%;  Pred. No. 2;
  Matches   87;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
              ||||| |    |||   |   | |   |  ||||||| || ||||   | |  |  |||
Db        370 CAGACCCAGCATGAAGAGCTCATCTGCGATGGCCAGGTTGAGGATGTAAATGTTGGTGAT 311

Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
                 |||  |||  |     |  ||| ||  |  ||| ||||     || | | |  ||
Db        310 GGTCTTCATCTTGGCATAGCGGAGGATGACATAAATGACAAGTGTGTTGCCACACAACCC 251

Qy        334 GATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393
               ||  | | | |  | ||   ||| |    |    | |  ||| |   || |  || | |
Db        250 AATGATGCAGACCACAAAATAGATGAATGTGAGGACTGCATTGCTTGTCAGGTCATAGTA 191

Qy        394 C 394
              |
Db        190 C 190


RESULT 15
US-09-134-000C-1552/c
; Sequence 1552, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE:  1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1552
;  LENGTH: 1797
;  TYPE: DNA
;  ORGANISM: Enterococcus faecalis
US-09-134-000C-1552

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 1797;
  Best Local Similarity   65.2%;  Pred. No. 2.3;
  Matches   45;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||||||   | | |  ||   |||||||| ||||||  |||||  |||  | | | | |
Db        721 CTGGTTTTTCTTCTGCTTCTTCATGAAATTGCACATATAAGCGCAATAATTCATTAATTG 662

Qy        421 GGTTTTGGC 429
              | | |||||
Db        661 GTTGTTGGC 653


Search completed: March 18, 2004, 20:20:31
Job time : 57.4971 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		34.4	7.5	2751	15	US-10-398-221-3563	Sequence 3563, Ap
c 2		34.2	7.5	616	12	US-10-424-599-88945	Sequence 88945, A
c 3		34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c 4		34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c 5		34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c 6		34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c 7		34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
c 8		34	7.5	1625	15	US-10-369-493-26980	Sequence 26980, A
c 9		33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
10		33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
11		33	7.2	7766	12	US-10-222-566-3	Sequence 3, Appli
12		33	7.2	7766	14	US-10-222-162-3	Sequence 3, Appli
13		33	7.2	7766	14	US-10-143-024-3	Sequence 3, Appli
14		33	7.2	368004	9	US-09-949-654-3	Sequence 3, Appli
15		32.8	7.2	1776	12	US-10-282-122A-24385	Sequence 24385, A
c 16		32.6	7.1	596	14	US-10-184-644-310	Sequence 310, App
c 17		32.6	7.1	596	14	US-10-184-634-310	Sequence 310, App
c 18		32.6	7.1	596	14	US-10-063-685-100	Sequence 100, App
19		32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
c 20		32.2	7.1	708	12	US-10-282-122A-27953	Sequence 27953, A
21		32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c 22		32	7.0	983	9	US-09-764-864-465	Sequence 465, App
23		31.8	7.0	750	14	US-10-184-644-104	Sequence 104, App
24		31.8	7.0	750	14	US-10-184-634-104	Sequence 104, App
c 25		31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
26		31.8	7.0	4589	10	US-09-764-891-8845	Sequence 8845, Ap
27		31.8	7.0	4590	10	US-09-764-891-8846	Sequence 8846, Ap
28		31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
29		31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
c 30		31.6	6.9	4830	12	US-10-282-122A-27869	Sequence 27869, A
c 31		31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
32		31.4	6.9	269	9	US-09-923-876-5903	Sequence 5903, Ap
33		31.4	6.9	269	11	US-09-923-876-5903	Sequence 5903, Ap
c 34		31.4	6.9	1287	15	US-10-369-493-46950	Sequence 46950, A
35		31.4	6.9	2007	12	US-10-425-114-31950	Sequence 31950, A
36		31.4	6.9	2074	12	US-10-425-114-27886	Sequence 27886, A
c 37		31	6.8	487	14	US-10-060-036-4434	Sequence 4434, Ap
38		30.8	6.8	2886	9	US-09-801-368-131	Sequence 131, App
c 39		30.6	6.7	421	10	US-09-918-995-33997	Sequence 33997, A
c 40		30.6	6.7	1110	11	US-09-826-509-570	Sequence 570, App
c 41		30.6	6.7	1110	14	US-10-225-567A-313	Sequence 313, App
42		30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
c 43		30.6	6.7	1351	15	US-10-305-720-1303	Sequence 1303, Ap
c 44		30.6	6.7	1510	15	US-10-292-798-677	Sequence 677, App
c 45		30.6	6.7	1608	9	US-09-815-242-3923	Sequence 3923, Ap

ALIGNMENTS

RESULT 1
US-10-398-221-3563
; Sequence 3563, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3563
; LENGTH: 2751
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3563

Query Match 7.5%; Score 34.4; DB 15; Length 2751;
Best Local Similarity 52.0%; Pred. No. 1.1;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy      238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
            |||||||   | |  ||| | |||| ||||||  ||      |       | || | | ||
Db      321 GATGTAGTGAAAGAAGCGAAAAATATTACTATCGACAAACCGTTTCCTCGTATGACATAT 380

Qy      298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGAT 357
            || |   | || || |  |     |  | | ||  | ||||||  | | ||   | |
Db      381 AAAGAAGCAATGGATCGTTTTGGCAGTGATAAACCAGATATTCGTTTTGGTTTAGAATTA 440
```

```
Qy        358 CAGCTGGTTACCCGTGTTGTTACCCATG 385
              ||    ||  ||  ||| || |   |||
Db        441 CAAAATGTATCCGATGTCGTAAAAGATG 468
```

RESULT 2
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(616)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

Query Match 7.5%; Score 34.2; DB 12; Length 616;
Best Local Similarity 50.9%; Pred. No. 0.56;
Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
              | | |  |||  ||  || ||| || || |||  | |    | |  | | | |    | |
Db        556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497

Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              | | ||   || |  | |     || ||  |  |    | | |||    ||   |  | |
Db        496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437

Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              || ||||||  |  |   ||| |||  ||||| ||| ||
Db        436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTA 398
```

RESULT 3
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |   |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 4
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |   |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 5
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |    |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 6
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-010-390-6

Query Match 7.5%; Score 34; DB 14; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |    |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 7
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (EBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-441-736-5

Query Match 7.5%; Score 34; DB 15; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |    |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 8
US-10-369-493-26980/c
; Sequence 26980, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 26980
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-26980

Query Match 7.5%; Score 34; DB 15; Length 1625;
Best Local Similarity 61.1%; Pred. No. 1.1;
Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```
Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              || |||||  || ||   ||| |||  | || |    || ||  |  || || ||||||
Db        395 TCGGAAATTCTGTCAAGCTCGGCCACAGCAATGTCGCAAGCTTGGTCATAACCGTCGGCG 336

Qy        350 AATATGATCAGCTGGTTACCCGTGTTGTTA 379
              | | ||||  | ||| | ||| |||   ||
Db        335 ATTCTGATGGGGTGGATGCCCTTGTCAATA 306
```

RESULT 9
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

```
  ;  APPLICANT: OMURA, SATOSHI
  ;  APPLICANT:  IKEDA, HARUO
  ;  APPLICANT:  ISHIKAWA, JUN
  ;  APPLICANT:  HORIKAWA, HIROSHI
  ;  APPLICANT:  SHIBA, TADAYOSHI
  ;  APPLICANT:  SAKAKI, YOSHIYUKI
  ;  APPLICANT:  HATTORI, MASAHIRA
  ;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
  ;  FILE REFERENCE: 249-262
  ;  CURRENT APPLICATION NUMBER: US/10/156,761
  ;  CURRENT FILING DATE:  2002-05-29
  ;  PRIOR APPLICATION NUMBER: JP 2001-204089
  ;  PRIOR FILING DATE: 2001-05-30
  ;  PRIOR APPLICATION NUMBER: JP 2001-272697
  ;  PRIOR FILING DATE: 2001-08-02
  ;  NUMBER OF SEQ ID NOS: 15109
  ; SEQ ID NO 3483
  ;    LENGTH: 1662
  ;    TYPE: DNA
  ;    ORGANISM: Streptomyces avermitilis
  ;    FEATURE:
  ;    NAME/KEY: CDS
  ;    LOCATION: (1)..(1662)
 US-10-156-761-3483

   Query Match             7.3%;  Score 33.2;  DB 14;  Length 1662;
   Best Local Similarity   52.1%;  Pred. No. 2.1;
   Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
              | || | ||  | ||||| | || |  | | || || || | ||| | |||||  |
Db        590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531

Qy        110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              |  ||| ||  | | |    ||   ||  | |     | || |||    | | |||
Db        530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

Qy        170 TTGCTCTGCAAAGCGATGCCCG 191
              |||  ||  |   |||  | | |
Db        470 TTGAACTTCTTGGCGGCGACGG 449


RESULT 10
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;    LENGTH: 9025608
;    TYPE: DNA
;    ORGANISM: Streptomyces avermitilis
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (4187715)
;    OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

   Query Match             7.3%;  Score 33.2;  DB 14;  Length 9025608;
   Best Local Similarity   52.1%;  Pred. No. 1.8e+02;
   Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
              | || | ||  | ||||| | || |  | | || || || | ||| | |||||  |
Db    4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

Qy        110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              |  ||| ||  | | |    ||   ||  | |     | || |||    | | |||
Db    4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

Qy        170 TTGCTCTGCAAAGCGATGCCCG 191
              |||  ||  |   |||  | | |
Db    4331901 TTGAACTTCTTGGCGGCGACGG 4331922


RESULT 11
US-10-222-566-3
; Sequence 3, Application US/10222566
; Publication No. US20040044192A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD3
;  CURRENT APPLICATION NUMBER: US/10/222,566
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;    LENGTH: 7766
;    TYPE: DNA
;    ORGANISM: Borrelia burgdorferi
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (127)
;    OTHER INFORMATION: R = A OR G
US-10-222-566-3

   Query Match             7.2%;  Score 33;  DB 12;  Length 7766;
   Best Local Similarity   47.8%;  Pred. No. 5.8;
   Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  |||||| ||||     |    ||||      ||  ||| |  | | | |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||     | |   |  | |||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |  || ||| |   |    | ||| |          | |||  |||||  |    |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATATGAT 357
              || ||||      |||||
Db       4612 TTTGCTGTGAAGAGTAATGAT 4632


RESULT 12
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
```

```
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD4
;  CURRENT APPLICATION NUMBER: US/10/222,162
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-162-3

  Query Match             7.2%;  Score 33;  DB 14;  Length 7766;
  Best Local Similarity   47.8%;  Pred. No. 5.8;
  Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |     ||||      ||  ||| |  | | |   | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |    |   | |||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    |  ||| |              | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATATGAT 357
               || ||||        |||||
Db       4612 TTTGCTGTGAAGAGTAATGAT 4632


RESULT 13
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD1
;  CURRENT APPLICATION NUMBER: US/10/143,024
;  CURRENT FILING DATE:  2002-05-10
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-143-024-3

  Query Match             7.2%;  Score 33;  DB 14;  Length 7766;
  Best Local Similarity   47.8%;  Pred. No. 5.8;
  Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |     ||||      ||  ||| |  | | |   | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |    |   | |||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    |  ||| |              | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATATGAT 357
               || ||||        |||||
Db       4612 TTTGCTGTGAAGAGTAATGAT 4632


RESULT 14
US-09-949-654-3
; Sequence 3, Application US/09949654
; Patent No. US20020127644A1
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000817
;  CURRENT APPLICATION NUMBER: US/09/949,654
;  CURRENT FILING DATE:  2001-09-12
;  PRIOR APPLICATION NUMBER: 60/231,572
;  PRIOR FILING DATE: 2000-09-11
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 368004
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(368004)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-654-3

  Query Match             7.2%;  Score 33;  DB 9;  Length 368004;
  Best Local Similarity   53.5%;  Pred. No. 48;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        328 AACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAA 387
              | ||| | | |||||| | || ||| |||||  | |  ||        | || |||| | |
Db     165120 ATCTCTGTTCTTACTGCCAGCAAATCTGATCTACAGAGTATAACATGTATTTCCCAAGCA 165179

Qy        388 ATGGCACATGCAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | || |   ||  | || ||      |  |||   | || || |  |  ||  |||
Db     165180 CAGACAAAGCAAAAGGAAAAGGGCTAGGCAGGAGGTAGGGACACCAGCTGCTGGGGAAAC 165239

Qy        448 CAGTATTAA 456
               ||| || |
Db     165240 AAGTCTTCA 165248


RESULT 15
US-10-282-122A-24385
; Sequence 24385, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24385
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Listeria monocytogenes
US-10-282-122A-24385

Query Match 7.2%; Score 32.8; DB 12; Length 1776;
Best Local Similarity 51.4%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||| | | | |  ||  | |||| || |||||||     |       | || | | | ||
Db        775 GATGTTGTCAAAGAAGCAAAAAATATCACCATTGAAAAACCGTTTCCTCGTATGACATAT 834

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGAT 357
              || |   | || || | |     | | | ||  | ||||||  | | ||   | |
Db        835 AAAGAAGCAATGGATCGTTTTGGCAGTGATAAACCGGATATTCGTTTTGGTTTAGAATTA 894

Qy        358 CAGCTGGTTACCCGTGTTGTTACCCATG 385
              ||    ||  ||  ||| || |   |||
Db        895 CAAAATGTATCCGATGTCGTAAAAGATG 922
```

Search completed: March 19, 2004, 03:49:29
Job time : 264.734 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-13	Sequence 13, Appl
2	404.8	88.8	456	6	US-08-233-642A-56	Sequence 56, Appl
3	404.8	88.8	456	23	US-09-543-407-1	Sequence 1, Appli
4	382.4	83.9	456	23	US-09-543-407-23	Sequence 23, Appl
5	376	82.5	456	23	US-09-543-407-11	Sequence 11, Appl
6	361.6	79.3	456	23	US-09-543-407-19	Sequence 19, Appl
7	355.2	77.9	456	23	US-09-543-407-17	Sequence 17, Appl
8	353.6	77.5	456	23	US-09-543-407-25	Sequence 25, Appl
9	352	77.2	456	23	US-09-543-407-15	Sequence 15, Appl
10	350.4	76.8	456	23	US-09-543-407-21	Sequence 21, Appl
11	350.4	76.8	456	23	US-09-543-407-27	Sequence 27, Appl
12	348.8	76.5	456	23	US-09-543-407-29	Sequence 29, Appl
13	287.2	63.0	361	6	US-08-233-642A-54	Sequence 54, Appl
14	227.2	49.8	477	17	US-09-252-691-172	Sequence 172, App
15	227.2	49.8	477	17	US-09-252-691C-172	Sequence 172, App
16	227.2	49.8	477	51	US-10-417-886-172	Sequence 172, App
17	225.6	49.5	456	23	US-09-543-407-3	Sequence 3, Appli
18	225.6	49.5	675	13	US-08-978-878-3	Sequence 3, Appli
19	225.6	49.5	675	31	US-09-741-873B-3	Sequence 3, Appli
20	223.2	48.9	453	92	US-60-352-946-1	Sequence 3, Appli
21	223.2	48.9	453	101	US-60-444-371-1	Sequence 1, Appli
22	181.6	39.8	396	13	US-08-978-878-1	Sequence 1, Appli
23	156.8	34.4	360	31	US-09-741-873B-1	Sequence 1, Appli
24	49	10.7	78	23	US-09-543-407-53	Sequence 1, Appli
c 25	48.4	10.6	78	23	US-09-543-407-52	Sequence 53, Appl
26	48	10.5	48	23	US-09-543-407-9	Sequence 52, Appl
c 27	44.8	9.8	2000	36	US-09-887-272A-5263	Sequence 9, Appli
c 28	41.6	9.1	78	23	US-09-543-407-48	Sequence 5263, Ap
29	40.8	8.9	100	45	US-10-146-492B-79	Sequence 48, Appl
30	40.6	8.9	78	23	US-09-543-407-49	Sequence 79, Appl
31	39.4	8.6	2732	51	US-10-437-963-37426	Sequence 49, Appl
32	37.6	8.2	456	23	US-09-543-407-2	Sequence 37426, A
33	37.6	8.2	522	17	US-09-252-691-170	Sequence 2, Appli
34	37.6	8.2	522	51	US-10-417-886-170	Sequence 170, App
35	37.6	8.2	522	23	US-09-543-407-4	Sequence 170, App
36	37.4	8.2	1304	51	US-10-437-963-19773	Sequence 170, App
37	37.4	8.2	100	45	US-10-146-492B-77	Sequence 4, Appli
38	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 19773, A
c 39	35.4	7.8	3412	1	PCT-US01-08631-24948	Sequence 77, Appl
c 40	35.4	7.8	100	45	US-10-146-492B-78	Sequence 24330, A
41	35	7.7	1083	1	PCT-US01-08631-12549	Sequence 24948, A
42	34.8	7.6	2751	50	US-10-398-221-3563	Sequence 78, Appl
43	34.4	7.5	385	18	US-09-333-535A-288	Sequence 12549, A
c 44	34.2	7.5	385	25	US-09-615-606A-909	Sequence 3563, Ap
c 45	34.2	7.5				Sequence 288, App
						Sequence 909, App

ALIGNMENTS

RESULT 1
US-09-543-407-13
, Sequence 13, Application US/09543407
, GENERAL INFORMATION:
, APPLICANT: White, Aaron P.
, APPLICANT: Doran, James L.
, APPLICANT: Collinson, S. Karen
, APPLICANT: Kay, William W.
, TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
, TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
, FILE REFERENCE: 920043.406
, CURRENT APPLICATION NUMBER: US/09/543,407
, CURRENT FILING DATE: 2000-04-05
, NUMBER OF SEQ ID NOS: 59
, SOFTWARE: FastSEQ for Windows Version 4.0
, SEQ ID NO 13
, LENGTH: 456
, TYPE: DNA
, ORGANISM: Artificial Sequence
;
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match
Best Local Similarity 100.0%; Score 456; DB 23; Length 456;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES:
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION
; NAME King, Joshua

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.9e-118;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||     | ||||  ||          |  || | ||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.9e-118;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||     | ||||  ||          |  || | ||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

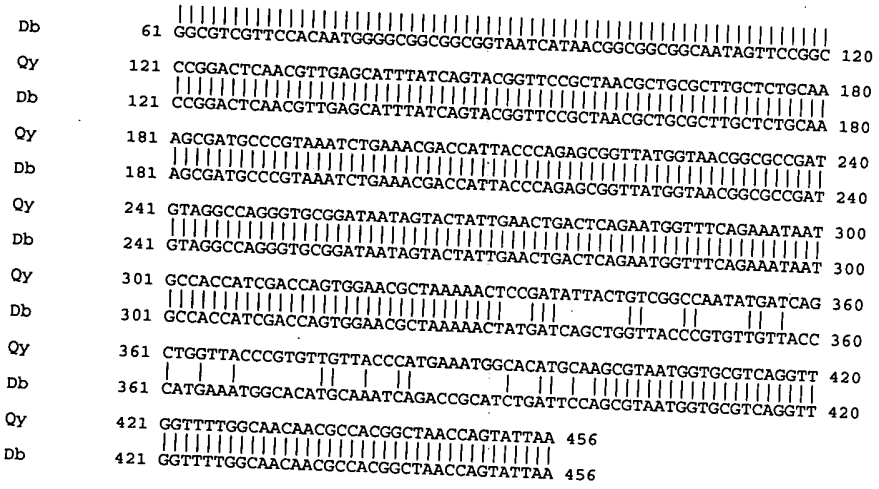
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             83.9%;  Score 382.4;  DB 23;  Length 456;
  Best Local Similarity   89.9%;  Pred. No. 2.8e-111;
  Matches  410;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

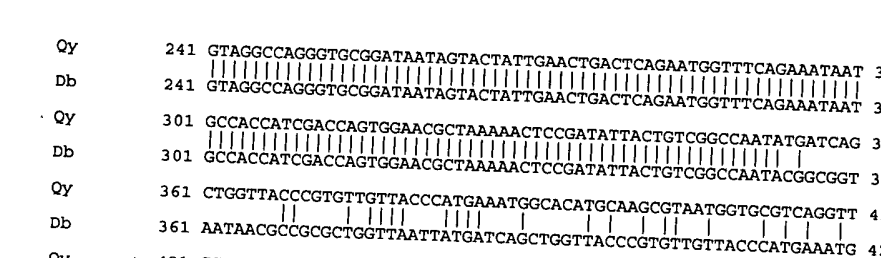
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```



RESULT 5
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

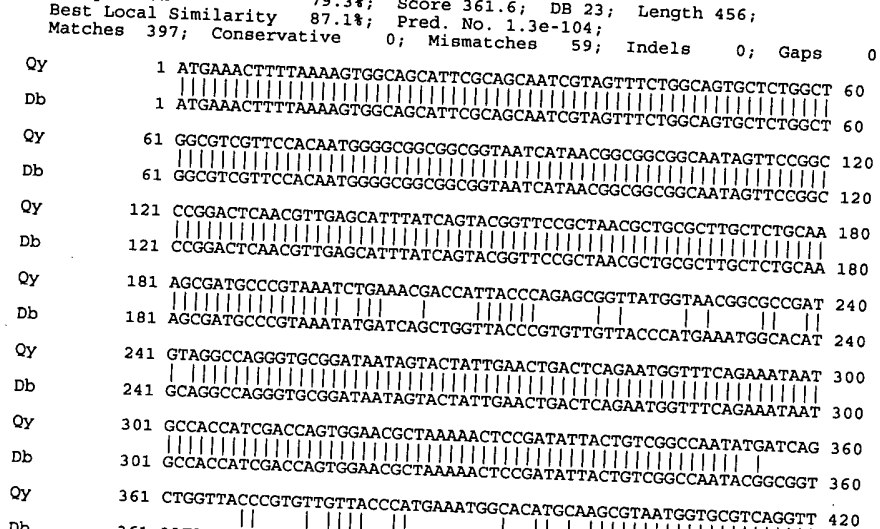
Query Match 82.5%; Score 376; DB 23; Length 456;
Best Local Similarity 89.0%; Pred. No. 3.2e-109;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;





RESULT 6
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 1.3e-104;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;



```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 1.5e-102;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||     |    ||   |    | | ||  | |     || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 4.8e-102;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

; LENGTH 456
; TYPE DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-15

Query Match
Best Local Similarity 77.2%; Score 352; DB 23; Length 456;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db 61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy 301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Db 301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Qy 361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
Db 361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy 421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db 421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-21

Query Match
Best Local Similarity 76.8%; Score 350.4; DB 23; Length 456;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy 241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATAT 300
Db 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy 301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Db 301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Qy 361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
Db 361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy 421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db 421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-27

Query Match
Best Local Similarity 76.8%; Score 350.4; DB 23; Length 456;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db 241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300
Qy 301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Db 301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy 361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
Db 361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy 421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db 421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||  |   |  |  ||  |   |     |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||   | ||||  ||        |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 76.5%; Score 348.8; DB 23; Length 456;
Best Local Similarity 85.3%; Pred. No. 1.7e-100;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||  || |||  |   |     |  |  |  |  |  |  |  |  |  |     |
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||   | ||||  ||        |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 63.0%; Score 287.2; DB 6; Length 361;
Best Local Similarity 99.0%; Pred. No. 8.7e-81;
Matches 289; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
              |||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292


RESULT 14
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match             49.8%;  Score 227.2;  DB 17;  Length 477;
  Best Local Similarity   68.6%;  Pred. No. 1.6e-61;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||  ||||||||||    | |||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||     | | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  ||||||||| ||||||||   || |||| | |||||||||  ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | || |||||  | ||||||  |||||||||||||   |||  ||||||||||||  |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||| |||||    ||| |||| | ||| |  ||    || | |
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  |   ||   |||  |  | |     ||| ||  | ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||||||||||| || ||  |||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


RESULT 15
US-09-252-691C-172
; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-172

  Query Match             49.8%;  Score 227.2;  DB 17;  Length 477;
  Best Local Similarity   68.6%;  Pred. No. 1.6e-61;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||  ||||||||||    | |||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||     | | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  ||||||||| ||||||||   || |||| | |||||||||  ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | || |||||  | ||||||  |||||||||||||   |||  ||||||||||||  |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||| |||||    ||| |||| | ||| |  ||    || | |
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  |   ||   |||  |  | |     ||| ||  | ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||||||||||| || ||  |||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


Search completed: March 19, 2004, 03:01:49
Job time : 2650.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	225.6	49.5	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	181.6	39.8	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33	7.2	3990	5	US-09-830-230A-629	Sequence 629, App
	4	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
	5	31.4	6.9	1588	6	US-10-767-701-14522	Sequence 14522, A
	6	30.6	6.7	7096	6	US-10-021-698A-892	Sequence 892, App
	7	30.6	6.7	196083	6	US-10-021-698A-707	Sequence 707, App
c	8	30.2	6.6	2007	6	US-10-775-972-153	Sequence 153, App
c	9	30.2	6.6	2148	6	US-10-775-972-154	Sequence 154, App
	10	30	6.6	1355	6	US-10-045-674A-593	Sequence 593, App
	11	29.6	6.5	2159	6	US-10-767-701-13730	Sequence 13730, A
c	12	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
	13	29.6	6.5	318752	6	US-10-767-471-10553	Sequence 10553, A
	14	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
c	15	29.2	6.4	1790242	6	US-10-767-471-10805	Sequence 10805, A
	16	29	6.4	670	6	US-10-767-701-15726	Sequence 15726, A
c	17	29	6.4	861	6	US-10-767-701-3616	Sequence 3616, Ap
c	18	28.8	6.3	1857	1	PCT-US04-05654-2371	Sequence 2371, Ap
	19	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	20	28.6	6.3	4634	6	US-10-791-799-10	Sequence 10, Appl
	21	28.4	6.2	1543	6	US-10-767-795-1782	Sequence 1782, Ap
	22	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	23	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	24	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
	25	28.2	6.2	798	6	US-10-100-683-2192	Sequence 2192, Ap
c	26	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
	27	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
	28	28	6.1	714	6	US-10-779-543-23469	Sequence 23469, A
c	29	27.8	6.1	390	5	US-09-796-692B-3647	Sequence 3647, Ap
	30	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
c	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	32	27.6	6.1	402	5	US-09-969-034-1464	Sequence 1464, Ap
c	33	27.6	6.1	564	6	US-10-767-701-16629	Sequence 16629, A
	34	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	35	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	36	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	37	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
	38	27.6	6.1	10251	6	US-10-045-674A-582	Sequence 582, App
c	39	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
	40	27.4	6.0	819	6	US-10-417-884A-2244	Sequence 2244, Ap
	41	27.4	6.0	823	6	US-10-767-701-20007	Sequence 20007, A
c	42	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
	43	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
	44	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
	45	27.2	6.0	547	6	US-10-767-701-24545	Sequence 24545, A

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 49.5%; Score 225.6; DB 5; Length 675;
Best Local Similarity 68.4%; Pred. No. 1.5e-66;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||| |||||| | ||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || ||||||||||  ||||| || || || || ||    || |  | |
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||  |||  |||       |  |  |   ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873C
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873C-1

  Query Match             39.8%;  Score 181.6;  DB 5;  Length 396;
  Best Local Similarity   66.2%;  Pred. No. 8.7e-52;
  Matches  262;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||||||    || ||||| |||| |  ||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||   || |||| |||| ||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
```

```
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || ||||||||||  ||||| || || || || ||    || |  | |
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||  |||  |||       |  |  |   ||| |  |||  |||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396


RESULT 3
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Lyme Disease Vaccines
;  FILE REFERENCE: PB481US
;  CURRENT APPLICATION NUMBER: US/09/830,230A
;  CURRENT FILING DATE:  2001-09-27
;  PRIOR APPLICATION NUMBER: PCT/US98/12718
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/057,483
;  PRIOR FILING DATE: 1997-09-03
;  PRIOR APPLICATION NUMBER: 60/053,344
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/053,377
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/050,359
;  PRIOR FILING DATE: 1997-06-20
;  NUMBER OF SEQ ID NOS: 756
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
;   LENGTH: 3990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1139)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1143)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1148)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1210)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1244)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1247)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1250)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1251)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629
```

```
  Query Match             7.2%;  Score 33;  DB 5;  Length 3990;
  Best Local Similarity   47.8%;  Pred. No. 0.45;
  Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  |||||  ||||    |    ||||      ||  ||| |  | | |  |  | |
Db       3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |   |  | |||||||||  | |||    |||  ||
Db       3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |   |  ||| |        | |||  |||||  |    |
Db       3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 3612

Qy        337 ATTACTGTCGGCCAATATGAT 357
               || ||||       |||||
Db       3613 TTTGCTGTGAAGAGTAATGAT 3633


RESULT 4
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
;  APPLICANT: Neubauer, Antonie
;  APPLICANT:  Ziegler, Christina
;  TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
;  FILE REFERENCE: 1/1372
;  CURRENT APPLICATION NUMBER: US/10/624,149A
;  CURRENT FILING DATE:  2003-07-21
;  PRIOR APPLICATION NUMBER: 60/403,282
;  PRIOR FILING DATE: 2002-08-14
;  PRIOR APPLICATION NUMBER: DE 10233064
;  PRIOR FILING DATE: 2002-07-19
;  PRIOR APPLICATION NUMBER: DE 10317008
;  PRIOR FILING DATE: 2003-04-11
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
;   LENGTH: 150223
;   TYPE: DNA
;   ORGANISM: Equine herpesvirus 1
;  PUBLICATION INFORMATION:
;   AUTHORS: Telford,E.A.
;   AUTHORS: Watson, M.S.
;   AUTHORS: McBride, K.
;   AUTHORS: Davison, A.J.
;   TITLE: The DNA sequence of equine herpesvirus-1
;   JOURNAL: Virology
;   VOLUME: 189
;   ISSUE: 1
;   PAGES: 304-316
;   DATE: JUL-1992
;   DATABASE ACCESSION NUMBER: NC 001491, NCBI
;   DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 150223;
  Best Local Similarity   51.4%;  Pred. No. 3.8;
  Matches   75;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              || ||  ||  |  | || ||  | ||| ||| |   |  ||| |||   | ||   |||
Db      95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy        233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
              |       |  |||| ||  ||||||    | ||  |  | | ||| |   |  |||
Db      95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy        293 GAAATAATGCCACCATCGACCAGTGG 318
              || |   |     || ||| | || |
Db      95662 GAGAGCTTATTCACAACGAGCTGTTG 95687


RESULT 5
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
;   LENGTH: 1588
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 1588;
  Best Local Similarity   44.5%;  Pred. No. 1;
  Matches  125;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
              | |  |||| |     | ||| | |  |      | ||||| |||  |  |  |  |
Db        455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
              |||||  |||  ||  ||   |||   |     |  |   |||           ||||  |  |
Db        515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy        160 AACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
                | ||   ||| |     |  | | |||  ||  | |||  |    | | ||  |     |
Db        575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              |||  |   | |  |||        |   |||||  || | |  ||  |  ||  |    ||
Db        635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAA 320
                | | ||  |   |  | ||||   || | ||  | |
Db        695 GCTGGTAGCTTTGGCGGTGGTTCCACTGGCGCCGAGAAGGA 735


RESULT 6
US-10-021-698A-892
; Sequence 892, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 892
;   LENGTH: 7096
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (7085)..(7087)
;   OTHER INFORMATION: a, t, c or g
US-10-021-698A-892

  Query Match             6.7%;  Score 30.6;  DB 6;  Length 7096;
  Best Local Similarity   48.6%;  Pred. No. 3.8;
  Matches   84;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        278 CTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATA 337
              |||   | ||     |||| ||   |   |   | |||     || | ||||  |  ||
Db       2964 CTCTCCAAGGGCCAGGAAACAACATCTTAAGATAAAAGTCACTGGCCATAAACAACCTTA 3023

Qy        338 TTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 397
              ||   |      ||| |  || || ||  ||| ||| ||| |    ||| |       |
Db       3024 TTCAAGGAAAATAATCTCGTCTGCAGGGCACCTGTGGTGTGAAGGCTGAGACCTGTGAGT 3083

Qy        398 CAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
                |||   |   |  | || || ||   ||     ||| || ||||| ||| |
Db       3084 TCAGCAACAGAATCGGCCAAGTGGGGGCTGCAGTAAACCCCTCGGCTTACCGG 3136


RESULT 7
US-10-021-698A-707
; Sequence 707, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 707
;   LENGTH: 196083
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (256)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (89496)..(89595)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (141214)..(141313)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (182468)..(182567)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (193506)..(193605)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (194799)..(194898)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (195354)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (195460)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (195644)
;   OTHER INFORMATION: a, t, c or g
US-10-021-698A-707

  Query Match             6.7%;  Score 30.6;  DB 6;  Length 196083;
  Best Local Similarity   48.6%;  Pred. No. 17;
  Matches   84;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        278 CTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATA 337
              |||   | ||     |||| ||   |   |   | |||     || | ||||  |  ||
Db      45914 CTCTCCAAGGGCCAGGAAACAACATCTTAAGATAAAAGTCACTGGCCATAAACAACCTTA 45973

Qy        338 TTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 397
              ||   |      ||| |  || || ||  ||| ||| ||| |    ||| |       |
Db      45974 TTCAAGGAAAATAATCTCGTCTGCAGGGCACCTGTGGTGTGAAGGCTGAGACCTGTGAGT 46033

Qy        398 CAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
                |||   |   |  | || || ||   ||     ||| || ||||| ||| |
Db      46034 TCAGCAACAGAATCGGCCAAGTGGGGGCTGCAGTAAACCCCTCGGCTTACCGG 46086


RESULT 8
US-10-775-972-153/c
; Sequence 153, Application US/10775972
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C21
;  CURRENT APPLICATION NUMBER: US/10/775,972
;  CURRENT FILING DATE:  2004-02-10
;  NUMBER OF SEQ ID NOS: 563
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-972-153

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 2007;
  Best Local Similarity   45.8%;  Pred. No. 2.9;
  Matches  104;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        140 TTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTG 199
              ||   ||| |   |  |   || | ||  | | | | |     | |  | ||||  |
Db       1416 TTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTAAGTTAAAAGTACTTAAAAGTT 1357

Qy        200 AAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              |  |   |||   ||  || ||| |  ||| |    |     |||| |   |  | ||
Db       1356 CACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCTGCCTAGGGGAATTTCAGCATT 1297

Qy        260 ATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGA 319
              |  |     |  ||| || ||| |     | |     |||||     ||  | |  |||||
Db       1296 AGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATAATCTGCCCTCCCATTAGTGGT 1237

Qy        320 ACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTT 366
              | |  | || |||| |  |    | |  |  || | | ||    |||
Db       1236 AGGAGAGAAGCTCCAAGTTCCGAGGCTTCTCATTTTACCAAACAGTT 1190


RESULT 9
```

```
US-10-775-972-154/c
; Sequence 154, Application US/10775972
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C21
;  CURRENT APPLICATION NUMBER: US/10/775,972
;  CURRENT FILING DATE:  2004-02-10
;  NUMBER OF SEQ ID NOS: 563
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 154
;   LENGTH: 2148
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-972-154

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 2148;
  Best Local Similarity   45.8%;  Pred. No. 3;
  Matches  104;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        140 TTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTG 199
              ||   ||| |   |  |   || | ||  | | | | |     | |  | ||||  |
Db       1549 TTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTAAGTTAAAAGTACTTAAAAGTT 1490

Qy        200 AAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              |  |   |||   ||  || ||| |  ||| |   |    |||| |   |  | ||
Db       1489 CACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCTGCCTAGGGGAATTTCAGCATT 1430

Qy        260 ATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGA 319
              |  |     |||  || ||| |   | |    |||||    ||  | |  |||||
Db       1429 AGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATAATCTGCCCTCCCATTAGTGGT 1370

Qy        320 ACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTT 366
              | |  | || |||| |  |    | |  |  || | | ||    |||
Db       1369 AGGAGAGAAGCTCCAAGTTCCGAGGCTTCTCATTTTACCAAACAGTT 1323


RESULT 10
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674A
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 06/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)

US-10-045-674A-593

  Query Match             6.6%;  Score 30;  DB 6;  Length 1355;
  Best Local Similarity   25.1%;  Pred. No. 2.8;
  Matches   84;  Conservative   80;  Mismatches  170;  Indels    0;  Gaps    0;

Qy         57 GGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTC 116
              :| :||:| :  :     : |  ::||:||:||:: : |   :|| ||: :: |  |:
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        117 CGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCT 176
              :||: : |       :   |:: :  : | | :| :| :| · ||    ||   :||:
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        177 GCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGC 236
              |: |:| :     : |:  :|         | : | | :    :| |:| :|| ||:
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        237 CGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAA 296
              :|   ::   | : | ||: |: || ||  :         :   | :||:  :   :
Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYATYGGYGAYGTYWS 1007

Qy        297 TAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              :  :    | | :| : |  |   : |:  : |:| :| :  : |:     :||:   |
Db       1008 YGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKAATTCYCARATGGC 1067

Qy        357 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 390
              :||: | | :  : |:| :  ::: | :  :|||
Db       1068 YCARGTYGGWGAYGGKGAYAAYWSWCCKYTWATG 1101


RESULT 11
US-10-767-701-13730
; Sequence 13730, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 13730
;   LENGTH: 2159
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS49140_1
US-10-767-701-13730

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 2159;
  Best Local Similarity   48.8%;  Pred. No. 4.8;
  Matches   80;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        288 TTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGG 347
              || ||  | | ||| ||||    |||  |    |  || |||||     ||||  | |   |
Db        682 TTACACTATTCATGACACCGAGGACACGGAAACTGCCAAAAAAGATGATACAAGTTATGA 741

Qy        348 CCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAAT 407
              |||  | |       |      ||| | |  |  ||   |  ||| |||   |||  ||
Db        742 CCAGCAGGGGAGCAAGCCCCTGTGTGATTTCTCAAATTTTAGAGCAAATGTGTGCGAGAT 801

Qy        408 GGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGT 451
              |    || | ||| |  ||  |   || || || |  | |  ||
Db        802 GCGTGGTGATGTTAGAGTTCACCCAAATGCTACAGAAATCATGT 845


RESULT 12
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
```

```
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
;   LENGTH: 24900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(24900)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 24900;
  Best Local Similarity   52.4%;  Pred. No. 15;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||  | || ||  | || | ||  |     |   | || || |  | | ||||| |
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |  |||||  |  | | |||||   | ||||| ||| | | | |      || |    |||
Db       9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929

Qy        327 AAAC 330
              |||
Db       9928 AATC 9925


RESULT 13
US-10-767-471-10553
; Sequence 10553, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10553
;   LENGTH: 318752
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(318752)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10553

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 318752;
  Best Local Similarity   52.4%;  Pred. No. 46;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              | |  | |  || |||||   || ||||||| ||     ||  ||  ||   | | | ||
Db     254839 TGTATACCACACCCAGAAACCTTGCAGAAATCATCATGACAGAGATGAGATAATCTCAAA 254898

Qy        327 AAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGA 386
              |   || ||| |||       |   || ||    ||||||      || || | |    | |
Db     254899 ACTTTCTGATTTTATGTGGCGATCATTTGCCATGCTGGTAGAAGGTTTTATGAGAACTTA 254958

Qy        387 AATG 390
              ||||
Db     254959 AATG 254962


RESULT 14
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 2.3;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||    |       |||| | |     | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
                 ||||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 15
US-10-767-471-10805/c
; Sequence 10805, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10805
;   LENGTH: 1790242
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1790242)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10805

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 1790242;
  Best Local Similarity   53.5%;  Pred. No. 1e+02;
  Matches   61;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        266 CTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              || ||    |||  |||||||   ||||  | | ||||   | |  ||    || | |
Db     369207 CTGTTCTGTTGATGCAGAATGAAATCAGGTAAACTGCCCTAAGCTTCCTCAAGATCTACT 369148

Qy        326 AAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTA 379
              || ||| | |  | ||||  ||| ||||       |  |||| | | || | |
Db     369147 CAAGCTCAGTTTATCCTGTTAGCCTATATCTCTTTTTCTTTACACATTTTTTAA 369094


Search completed: March 19, 2004, 03:10:11
Job time : 58.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	226.2	49.6	680	12	BJ618688	BJ618688 BJ618688
	2	36	7.9	523	12	BJ334624	BJ334624 BJ334624
c	3	35.6	7.8	744	14	CF446143	CF446143 EST682488
c	4	34.8	7.6	549	28	BZ295150	BZ295150 CG1320.r1
	5	34.8	7.6	594	12	BJ335653	BJ335653 BJ335653
	6	34.8	7.6	813	28	BZ204853	BZ204853 CH230-390
	7	34.6	7.6	540	14	CB210843	CB210843 OML01123
	8	34.6	7.6	556	9	AV976624	AV976624 AV976624
	9	34.4	7.5	539	13	BQ141266	BQ141266 NF017G03P
	10	34.4	7.5	560	10	AW686303	AW686303 NF036D06N
	11	34.4	7.5	627	10	AW573882	AW573882 EST316473
	12	34.2	7.5	456	12	BJ331339	BJ331339 BJ331339
	13	34.2	7.5	524	12	BJ367372	BJ367372 BJ367372
	14	34.2	7.5	547	12	BJ333958	BJ333958 BJ333958
	15	34.2	7.5	553	12	BJ336903	BJ336903 BJ336903
	16	34.2	7.5	569	12	BJ334047	BJ334047 BJ334047
	17	34.2	7.5	574	12	BJ330328	BJ330328 BJ330328
	18	34.2	7.5	607	12	BJ365766	BJ365766 BJ365766
	19	34.2	7.5	625	12	BJ396922	BJ396922 BJ396922
	20	34.2	7.5	629	12	BJ335378	BJ335378 BJ335378
	21	34.2	7.5	635	12	BJ397391	BJ397391 BJ397391
	22	34.2	7.5	645	12	BJ392753	BJ392753 BJ392753
	23	34.2	7.5	670	12	BJ336637	BJ336637 BJ336637
	24	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	25	34	7.5	234	10	BB581987	BB581987 BB581987
c	26	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	27	34	7.5	512	10	BB757100	BB757100 BB757100
	28	33.8	7.4	298	14	F19524	F19524 HSPD04128 H
c	29	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	30	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	31	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
	32	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	33	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
c	34	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	35	33.6	7.4	811	28	BZ469000	BZ469000 BOOAO58TF
	36	33.4	7.3	402	13	BU767674	BU767674 SJEBBF08
c	37	33.4	7.3	505	6	AL809250	Al809250 Triticum
	38	33.4	7.3	536	13	BU776718	BU776718 SJEDCA03
	39	33.4	7.3	546	13	BU780674	BU780674 SJEFZH02
c	40	33.4	7.3	699	12	BI959328	BI959328 HVSMEn001
	41	33.4	7.3	715	13	BU444683	BU444683 603765715
	42	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
	43	33.2	7.3	629	9	AU295325	AU295325 AU295325
	44	33.2	7.3	788	9	AU139321	AU139321 AU139321
c	45	33.2	7.3	1646	10	BF579120	BF579120 602096054

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 49.6%; Score 226.2; DB 12; Length 680;
Best Local Similarity 68.4%; Pred. No. 3.5e-53;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||    ||| |||||||||    ||     ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || ||||||||||  ||||| || || || || ||    || |  | |
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||  |||  |||        |  |  |  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
BJ334624
LOCUS BJ334624 523 bp mRNA linear EST 05-MAR-2002
DEFINITION BJ334624 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda47o06 5', mRNA sequence.
ACCESSION BJ334624
VERSION BJ334624.1 GI:19164754
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 523)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .523
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47o06"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Query Match 7.9%; Score 36; DB 12; Length 523;
Best Local Similarity 55.5%; Pred. No. 30;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||   ||  |||
Db        398 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGANGACAATAAT 457

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     |||||| |   |   |  ||||
Db        458 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAATAANGATGATGATGA 516
```

RESULT 3
CF446143/c
LOCUS CF446143 744 bp mRNA linear EST 04-SEP-2003
DEFINITION EST682488 normalized cDNA library of onion Allium cepa cDNA clone ACAHU74, mRNA sequence.
ACCESSION CF446143
VERSION CF446143.1 GI:34468845
KEYWORDS EST.
SOURCE Allium cepa (onion)
ORGANISM Allium cepa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae; Allium.
REFERENCE 1 (bases 1 to 744)
AUTHORS Havey,M.J., Cheung,F., Van Aken,S., Utterback,T. and Town,C.D.
TITLE Expressed Sequence Tags from a normalized library of mixed onion tissues (Allium cepa)
JOURNAL Unpublished (2003)
COMMENT Contact: Havey MJ
Department of Horticulture
USDA-ARS and University of Wisconsin
1575 Linden Drive, Madison, WI 53706, USA
Tel: 608-262-1830
Fax: 608-262-4743
Email: mjhavey@facstaff.wisc.edu
TIGR sequence name ACAHU74TR. For more information:
http://haveylab.hort.wisc.edu
Seq primer: CAG GAA ACA GCT ATG ACC.
FEATURES Location/Qualifiers
source 1. .744
/organism="Allium cepa"
/mol_type="mRNA"
/cultivar="Red Creole(bulbs), unknown(callus), Ebano & Texas Legend(roots)"
/db_xref="taxon:4679"
/clone="ACAHU74"
/tissue_type="Callus, roots, and young bulbs"
/clone_lib="normalized cDNA library of onion"
/note="Vector: pCMVSport6.1-ccdb (Invitrogen); Site_1: EcoRV (5'); Site_2: NotI (3'); Equal molar amounts of mRNA from callus, roots, and young bulbs were combined to synthesize the library. Normalization to enrich for low-copy transcripts was performed by proprietary techniques of Invitrogen."
ORIGIN

Query Match 7.8%; Score 35.6; DB 14; Length 744;
Best Local Similarity 47.7%; Pred. No. 46;

```
  Matches  104;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy        234 CGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAG 293
              ||  | || | || |      |||  | || |  | | ||||||   |  | | || |
Db        382 CGAAGCTGAAAGCAACAAACTTTATAGCAATATTGATCATCTGACTGCTATGGATATCCG 323

Qy        294 AAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATA 353
              ||||||    |  | |  | | |||| | ||||||  |  | |  |||| ||    | |
Db        322 AAATAACAATAGTAACAGCAAATGGATCTCTAAAAGGTGTGTTGATACTATCAATAATGA 263

Qy        354 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCG 413
               || | |   |  | |    | ||    |||||   |  |||      |  | || | |
Db        262 AGAGCGGAGAGAAAGCAACTCTTTTTATGATGAATCTGTTCATAATGTTGGCAAGGAGGG 203

Qy        414 TCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGT 451
                  ||       |||||||    ||||     |||
Db        202 AAGATTTAACAACTTCAACAACAAATCGGCCTCGAAGT 165


RESULT 4
BZ295150/c
LOCUS       BZ295150                549 bp    DNA     linear   GSS 31-OCT-2002
DEFINITION  CG1320.r1 Candida glabrata Random Genomic Library Candida glabrata
            genomic clone CG1320, genomic survey sequence.
ACCESSION   BZ295150
VERSION     BZ295150.1  GI:24436228
KEYWORDS    GSS.
SOURCE      Candida glabrata
  ORGANISM  Candida glabrata
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
  TITLE     Evidence from comparative genomics for a complete sexual cycle in
            the 'asexual' pathogenic yeast Candida glabrata
  JOURNAL   Genome Biol. 4 (2), R10 (2003)
  MEDLINE   22508158
   PUBMED   12620120
COMMENT     Contact: Wong S
            Department of Genetics, Smurfit Institute
            Trinity College Dublin
            Dublin 2, Ireland
            Tel: 353 1 6082319
            Fax: 353 1 6798558
            Email: swong@tcd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Candida glabrata"
                     /mol_type="genomic DNA"
                     /strain="CBS 138"
                     /db_xref="taxon:5478"
                     /clone="CG1320"
                     /clone_lib="Candida glabrata Random Genomic Library"
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 28;  Length 549;
  Best Local Similarity   52.0%;  Pred. No. 68;
  Matches   78;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        222 TTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCA 281
              |||| ||    || || |   ||||||   |     | | |  |     || |    ||
Db        291 TTATAGTTTTTGCTCCTACAAAGGCCAAATTTTCAGTGAAAAAAGCTGAAAATTTTTGCA 232

Qy        282 GAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTAC 341
              |||   |  |||     |  || ||   ||   | || ||| |  | |||| |||| |
Db        231 GAAAATTGACAGTGTGTGTATCAGCAAACACATCTAGATCGCAACCAGCTCCTATATCAG 172

Qy        342 TGTCGGCCAATATGATCAGCTGGTTACCCG 371
              |||| |||||  |  |||   | | || |||
Db        171 TGTCTGCCAAGAATATCCTTTTGATAACCG 142


RESULT 5
BJ335653
LOCUS       BJ335653                594 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ335653 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda51h06 5', mRNA sequence.
ACCESSION   BJ335653
VERSION     BJ335653.1  GI:19165783
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda51h06"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 12;  Length 594;
  Best Local Similarity   55.5%;  Pred. No. 70;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   |||| | | | |   ||||    ||  |||
Db        223 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 282

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     |||||| |   |   |  ||||
Db        283 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAATAAAGATGATGATGA 341


RESULT 6
BZ204853
LOCUS       BZ204853                813 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-390D21, genomic survey sequence.
ACCESSION   BZ204853
VERSION     BZ204853.1  GI:23862905
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-390D21.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
```

```
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 390  row: D  column: 21
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-390D21"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 28;  Length 813;
  Best Local Similarity   50.0%;  Pred. No. 80;
  Matches   87;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              ||| |   | | | || |    |   ||    ||  | || |  |   |    || |  ||
Db        548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
                || |   || ||       || | | |  |||||  |     |||||  ||  |   |
Db        608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAA 351
              ||   ||  |||||||| ||| |    | |||| | |   | |||| ||| ||
Db        668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAA 721


RESULT 7
CB210843
LOCUS       CB210843                540 bp    mRNA    linear   EST 05-FEB-2003
DEFINITION  OML01123 Oryza minuta HybriZAP-2.1 XR library Oryza minuta cDNA 5',
            mRNA sequence.
ACCESSION   CB210843
VERSION     CB210843.1  GI:28256934
KEYWORDS    EST.
SOURCE      Oryza minuta
  ORGANISM  Oryza minuta
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Shin,J.S.
  TITLE     Oryza minuta HybriZAP-2.1 XR library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Jeong Sheop Shin
            Plant Molecular Genetics
            Graduate School of Biotechnology, University of Korea
            136-701 Anam-dong 5/1 Seoul, Korea
            Tel: 00 82 2 3290 3430
            Fax: 00 82 2 927 9028
            Email: jsshin@kuccnx.korea.ac.kr.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Oryza minuta"
                     /mol_type="mRNA"
                     /db_xref="taxon:63629"
                     /dev_stage="4-weeks after germination"
                     /clone_lib="Oryza minuta HybriZAP-2.1 XR library"
                     /note="Organ: immature leaf"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 14;  Length 540;
  Best Local Similarity   50.3%;  Pred. No. 77;
  Matches   85;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        225 TGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAA 284
              |||   | | |    || |||    || |  ||||||  ||    |||||||  |
Db        140 TGGAGCCTGTGGTCCTGGAGGTTTTGGAGATGATAATGATAAATCTGAACTGTTTATTGT 199

Qy        285 TGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGT 344
              | |||||  ||  ||||| ||   |||    | ||| |||   ||  |  ||| |  | |
Db        200 TCGTTTCCAAAGAAATGCAACATACGAAGCATTGAAAGCTGCGAAAACAAATAATCTTAT 259

Qy        345 CGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393
              |       |    | |  |||   || ||   ||||   | |||||
Db        260 GGAGGCTCTCCAAAGACAGCATACTACTGATGCCTCCCAGCTAGTGGCA 308


RESULT 8
AV976624
LOCUS       AV976624                556 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  AV976624 Nori Satoh unpublished cDNA library, egg Ciona
            intestinalis cDNA clone cieg42j24 5', mRNA sequence.
ACCESSION   AV976624
VERSION     AV976624.1  GI:19466390
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cieg42j24"
                     /tissue_type="whole animal"
                     /dev_stage="egg"
                     /clone_lib="Nori Satoh unpublished cDNA library, egg"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 9;  Length 556;
  Best Local Similarity   47.1%;  Pred. No. 78;
  Matches  106;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy        222 TTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCA 281
              | |||  || |      |||||    |  | ||  |  |  | ||| |  ||     ||| |
Db         67 TGATGTCAAAGCTCATGATGTCACACCTGATGATGTCACCACTAATGATGTCACCACTAA 126

Qy        282 GAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTAC 341
                |||    ||  ||| ||| |||||   |  |     || ||| |  |  |  || ||
Db        127 TGATGTCACCACTAATGATGTCACCACTAATGATGTCAAAGCTGATGATGTCACTACTAA 186

Qy        342 TGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAG 401
              ||  | |   |  |||| |  |   |||   || ||| || | |||    |    | | |
Db        187 TGATGTCACACCTGATAATGTCACTACTAATGATGTCACACCTGATGATGTCACTACTAA 246

Qy        402 CGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAA 446
              |  |      | | | ||    |   ||  | | |||  ||||
```

```
Db        247 TGATGTCAAAGCTGATGATGTCACTACTAATGATGTCACACCTAA 291


RESULT 9
BQ141266
LOCUS       BQ141266                539 bp    mRNA    linear   EST 26-APR-2002
DEFINITION  NF017G03PH1F1024 Phoma-infected Medicago truncatula cDNA clone
            NF017G03PH 5', mRNA sequence.
ACCESSION   BQ141266
VERSION     BQ141266.1  GI:20277392
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
            Gonzales,R.A., Bell,C.J., Inman,J.T., Waugh,M.E., Sullivan,J.P.,
            May,G.D. and Paiva,N.L.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula Phoma-infected library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Paiva NL
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7317
            Fax: 580 221 7380
            Email: nlpaiva@noble.org
            Insert Length: 539   Std Error: 0.00
            Plate: 017  row: G  column: 03
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF017G03PH"
                     /tissue_type="leaf"
                     /dev_stage="Pathogen-induced, young trifoliate"
                     /clone_lib="Phoma-infected"
                     /note="Vector: pBluescript SK(-); Young trifoliate leaves
                     of Medicago truncatula were excised and dip-inoculated in
                     a spore suspension of Phoma medicaginis, and incubated in
                     humid dishes. Pools of leaves were harvested at 0, 15, and
                     30 minutes and 1, 2, 3, 6, 14, 24, 48, 72, and 96, hours,
                     and used to prepare total RNA. cDNA was prepared from
                     polyA+ enriched, pooled samples of equivalent amounts of
                     total RNA from each sample. The cDNA was directionally
                     ligated into the Uni-Zap XR vector (Stratagene) and
                     packaged using the Gigapack III Gold packaging extracts.
                     Phagemids containing cDNA inserts were in vivo excised
                     from the recombinant Uni-ZAP XR vector using ExAssist
                     helper phage and the E. coli strain XL1-Blue MRF'
                     (Stratagene). Excised plasmids were plated using SOLR
                     cells."
ORIGIN

  Query Match             7.5%;  Score 34.4;  DB 13;  Length 539;
  Best Local Similarity   46.0%;  Pred. No. 88;
  Matches  116;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy        129 AACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||  || |  | | | | | ||||| |  ||| ||    | |        ||  ||| ||
Db        233 AATTTTAAAAACATCTGACTCCGGTTACAATAATGCATACCCTTTGAATGAAGACGAGGC 292

Qy        189 CCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCA 248
              |||  |  |      ||    | | ||        |       | || | ||   |   |
Db        293 ACGTGATCCACCTGAGATTCCTCCACAATTGCAAAACACCTTGGTCGGCTATCCTGCTAA 352

Qy        249 GGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCAT 308
```

```
              |  |    | |||| || ||  | || |||||||  |   |   |||| ||       ||
Db        353 TGTCGATAGTTTTAGTTCTCTTCCATTGCCTCAGAACGCAATTCTAAATCATCTTTATAT 412

Qy        309 CGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTAC 368
              ||  |  |  |  |  |||  || |   |  || | ||  |   | |||   |   |||
Db        413 TGAGAATAGAGAGCCACAAAGATCTGTGGTGGCTTTGGGATACACTCATCGCTTTCGTAC 472

Qy        369 CCGTGTTGTTAC 380
              |    |||||||
Db        473 CAAATTTGTTAC 484


RESULT 10
AW686303
LOCUS       AW686303                560 bp    mRNA    linear   EST 20-DEC-2000
DEFINITION  NF036D06NR1F1000 Nodulated root Medicago truncatula cDNA clone
            NF036D06NR 5', mRNA sequence.
ACCESSION   AW686303
VERSION     AW686303.2  GI:11931130
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
            Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W.,
            May,G.D. and Paiva,N.L.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula nodulated root library
  JOURNAL   Unpublished (2000)
COMMENT     On Apr 14, 2000 this sequence version replaced gi:7561039.
            Contact: Paiva NL
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7317
            Fax: 580 221 7380
            Email: nlpaiva@noble.org
            Insert Length: 665   Std Error: 0.00
            Plate: 036  row: D  column: 06
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF036D06NR"
                     /tissue_type="root"
                     /dev_stage="Pooled developmental"
                     /clone_lib="Nodulated root"
                     /note="Vector: Lambda Zap; Four-week old Rhizobium
                     meliloti-inoculated Medicago truncatula roots, containing
                     a mixture of young and old roots and nodules."
ORIGIN

  Query Match             7.5%;  Score 34.4;  DB 10;  Length 560;
  Best Local Similarity   46.0%;  Pred. No. 89;
  Matches  116;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy        129 AACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||  || |  | | | | | ||||| |  ||| ||    | |        ||  ||| ||
Db        230 AATTTTAAAAACATCTGACTCCGGTTACAATAATGCATACCCTTTGAATGAAGACGAGGC 289

Qy        189 CCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCA 248
              |||  |  |      ||    | | ||        |       | || | ||   |   |
Db        290 ACGTGATCCACCTGAGATTCCTCCACAATTGCAAAACACCTTGGTCGGCTATCCTGCTAA 349

Qy        249 GGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCAT 308
              |  |    | |||| || ||  | || |||||||  |   |   |||| ||       ||
Db        350 TGTCGATAGTTTTAGTTCTCTTCCATTGCCTCAGAACGCAATTCTAAATCATCTTTATAT 409
```

```
Qy        309 CGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTAC 368
              ||  |  |  |  |  |||  || |  |  || | ||  |  | |||  |  |||
Db        410 TGAGAATAGAGAGCCACAAAGATCTGTGGTGGCTTTGGGATACACTCATCGCTTTCGTAC 469

Qy        369 CCGTGTTGTTAC 380
              |  |||||||
Db        470 CAAATTTGTTAC 481
```

RESULT 11
AW573882
LOCUS AW573882 627 bp mRNA linear EST 07-SEP-2000
DEFINITION EST316473 GVN Medicago truncatula cDNA clone pGVN-49K8, mRNA sequence.
ACCESSION AW573882
VERSION AW573882.1 GI:7238615
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 627)
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name:M252780e
TIGR sequence name:MTCAJ64TK
More information is available at.
'http://chrysie.tamu.edu/medicago'
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES Location/Qualifiers
source 1. .627
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-49K8"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XLOLR"
/clone_lib="GVN"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells."
ORIGIN

Query Match 7.5%; Score 34.4; DB 10; Length 627;
Best Local Similarity 46.0%; Pred. No. 93;
Matches 116; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy        129 AACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||  || |  |  | | | | |||||  |  ||| ||    | |        ||  ||| ||
Db        177 AATTTTAAAAACATCTGACTCCGGTTACAATAATGCATACCCTTTGAATGAAGACGAGGC 236

Qy        189 CCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCA 248
              |||  |  |  |  ||  | | ||  |  |  |  | ||  | |  |  |  |
Db        237 ACGTGATCCACCTGAGATTCCTCCACAATTGCAAAACACCTTGGTCGGCTATCCTGCTAA 296

Qy        249 GGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCAT 308
              |  |    |  |||| || ||  | || |||||||  |    |  ||||  ||      ||
Db        297 TGTCGATAGTTTTAGTTCTCTTCCATTGCCTCAGAACGCAATTCTAAATCATCTTTATAT 356

Qy        309 CGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTAC 368
              ||  |  |  |  |  |||  || |  |  || | ||  |  | |||  |  |||
Db        357 TGAGAATAGAGAGCCACAAAGATCTGTGGTGGCTTTGGGATACACTCATCGCTTTCGTAC 416

Qy        369 CCGTGTTGTTAC 380
              |  |||||||
Db        417 CAAATTTGTTAC 428
```

RESULT 12
BJ331339
LOCUS BJ331339 456 bp mRNA linear EST 05-MAR-2002
DEFINITION BJ331339 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda34m22 5', mRNA sequence.
ACCESSION BJ331339
VERSION BJ331339.1 GI:19161469
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 456)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .456
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda34m22"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 456;
Best Local Similarity 55.5%; Pred. No. 93;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||  ||   | |     ||| |||||   |||||  | | | |   ||||    ||  |||
Db        264 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 323

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  ||  ||  |  ||  |    ||   ||||||     ||||||  |   |   |  ||||
Db        324 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGATGA 382
```

RESULT 13
BJ367372
LOCUS BJ367372 524 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ367372 Dictyostelium discoideum cDNA library, CF Dictyostelium discoideum cDNA clone ddc42p08 5', mRNA sequence.
ACCESSION BJ367372
VERSION BJ367372.1 GI:19276674

```
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc42p08"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 524;
  Best Local Similarity   55.5%;  Pred. No. 99;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        181 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 240

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     ||||||  |   |   |  ||||
Db        241 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGATGA 299


RESULT 14
BJ333958
LOCUS       BJ333958                547 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ333958 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda44j16 5', mRNA sequence.
ACCESSION   BJ333958
VERSION     BJ333958.1  GI:19164088
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda44j16"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 547;
  Best Local Similarity   55.5%;  Pred. No. 1e+02;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        188 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 247

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     ||||||  |   |   |  ||||
Db        248 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGATGA 306


RESULT 15
BJ336903
LOCUS       BJ336903                553 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ336903 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda55g13 5', mRNA sequence.
ACCESSION   BJ336903
VERSION     BJ336903.1  GI:19167033
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda55g13"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 553;
  Best Local Similarity   55.5%;  Pred. No. 1e+02;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        184 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 243

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     ||||||  |   |   |  ||||
Db        244 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGATGA 302


Search completed: March 18, 2004, 20:11:27
Job time : 2035.55 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		403.2	88.4	456	6	I44909	I44909 Sequence 58
2		403.2	88.4	2067	1	SEU43280	U43280 Salmonella
3		395.2	86.7	5103	1	STAJ2301	AJ002301 Salmonell
4		395.2	86.7	22411	1	AE008749	AE008749 Salmonell
5		393.6	86.3	254050	1	AL627269	AL627269 Salmonell
6	c	393.6	86.3	301983	1	AE016840	AE016840 Salmonell
7		387.8	85.0	1048	1	STAGFBA	AJ000514 Salmonell
8		284.2	62.3	361	6	I44908	I44908 Sequence 56
9		278.4	61.1	2889	1	CSP515700	AJ515700 Citrobact
10		229.2	50.3	2920	1	CFR515701	AJ515701 Citrobact
11		225.4	49.4	1711	1	AF275733	AF275733 Escherich
12		225.4	49.4	10190	1	AE005315	AE005315 Escherich
13		225.4	49.4	327773	1	AP002554	AP002554 Escherich
14		224	49.1	4680	1	ECCSGABDG	X90754 E.coli csgG
15		224	49.1	10346	1	AE000205	AE000205 Escherich
16		224	49.1	15047	1	D90741	D90741 Escherichia
17		223.8	49.1	306358	1	AE016759	AE016759 Escherich
18		222.4	48.8	456	6	AX814811	AX814811 Sequence
19		222.4	48.8	648	1	ECOCSGAA	L04979 Escherichia
20		208	45.6	2883	1	ESA515702	AJ515702 Enterobac
21		203.2	44.6	230	1	SEU53207	U53207 Salmonella
22		186.4	40.9	10370	1	AE015131	AE015131 Shigella
23		186.4	40.9	292504	1	AE016981	AE016981 Shigella
24	c	163.4	35.8	437	1	AF237726	AF237726 Shigella
25		93.2	20.4	19201	1	D90742	D90742 Escherichia
26		50	11.0	1212	1	ECO131756	AJ131756 Escherich
27		48.2	10.6	78	6	AX814809	AX814809 Sequence
28		41.6	9.1	72	6	AX814798	AX814798 Sequence
29	c	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
30	c	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
31		35.6	7.8	13430	1	AE000409	AE000409 Escherich
32		35.6	7.8	108366	2	AC020660	AC020660 Homo sapi
33		35.6	7.8	110000	1	ECOUW67_2	Continuation (3 of
34	c	35.6	7.8	110000	2	AC145807_0	AC145807 Silurana
35	c	35.6	7.8	196539	2	AC145912	AC145912 Gallus ga
36	c	35.4	7.8	168267	2	AC146009	AC146009 Pan trogl
37		35.2	7.7	163619	2	AC119493	AC119493 Rattus no
38		34.6	7.6	11160	1	AE008209	AE008209 Agrobacte
39	c	34.6	7.6	11476	1	AE009405	AE009405 Agrobacte
40		34.6	7.6	125020	9	AF429315	AF429315 Homo sapi
41		34.2	7.5	1458	1	AF173043	AF173043 Orientia
42	c	34.2	7.5	179906	2	AC119683	AC119683 Rattus no
43		34.2	7.5	268098	2	AC133387	AC133387 Rattus no
44		34.2	7.5	299708	2	AC110963	AC110963 Rattus no
45		34	7.5	956	8	AK058943	AK058943 Oryza sat

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match              88.4%;  Score 403.2;  DB 6;  Length 456;
  Best Local Similarity    92.8%;  Pred. No. 2.7e-111;
  Matches  423;  Conservative.   0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match              88.4%;  Score 403.2;  DB 1;  Length 2067;
  Best Local Similarity    92.8%;  Pred. No. 3.1e-111;
```

```
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |    | |    |   ||   ||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match          86.7%;  Score 395.2;  DB 1;  Length 5103;
  Best Local Similarity  91.7%;  Pred. No. 9.1e-109;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |  ||  ||  | | | |  |  | |  |  |  ||  ||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  ||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421

RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
     RBS                complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene               complement(2580. .4088)
                        /gene="STM1128"
     CDS                complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS                complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene               complement(4413. .5099)
                        /gene="STM1129"
     CDS                complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
     RBS                complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene               5599. .6759
                        /gene="STM1130"
     CDS                5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVPLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene               6793. .7497
                        /gene="STM1131"
     RBS                6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           86.7%;  Score 395.2;  DB 1;  Length 22411;
  Best Local Similarity  91.7%;  Pred. No. 1e-108;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEIILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

/note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature"
gene 3001. .4011
/gene="STY1079"
/note="synonym: pyrD"
CDS 3001. .4011
/gene="STY1079"
/note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q8Z7S9"
/db_xref="SWISS-PROT:Q8Z7S9"
/translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
HI"
misc_feature 3136. .3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHOdehase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
misc_feature 3238. .3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
misc_feature 3874. .3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
gene 4174. .4719
/gene="STY1080"
CDS 4174. .4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepency in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q8Z7S8"
/translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
IMNDRLKPQVHCHSFSLEQAV"
gene complement(4716. .5825)
/gene="STY1081"
CDS complement(4716. .5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Fasta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q8Z7S7"
/db_xref="SPTREMBL:Q8Z7S7"
/translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
misc_feature complement(4737. .4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
gene 5924. .8032
/gene="STY1082"
CDS 5924. .8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q8Z7S6"
/db_xref="SPTREMBL:Q8Z7S6"
/translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
LSPDFARNRQIHNCWLIRAA"
misc_feature 5927. .7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, score 781.30, E-value 3.7e-231"
misc_feature 6506. .6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
gene 8045. .9952
/gene="STY1083"
CDS 8045. .9952
/gene="STY1083"
/note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa overlap"

Query Match 86.3%; Score 393.6; DB 1; Length 254050;
Best Local Similarity 91.4%; Pred. No. 3.9e-108;
Matches 417; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy 61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
Db 88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db 89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db 89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.3%;  Score 393.6;  DB 1;  Length 301983;
Best Local Similarity   91.4%;  Pred. No. 4e-108;
  Matches  417;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   |  | |     |   ||   |||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
     gene            KTAVVVQKQSHMAIRVTQR"
                     593. .1048
     CDS             /gene="agfA"
                     593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
     ORIGIN          IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"

       Query Match             85.0%;  Score 387.8;  DB 1;  Length 1048;
       Best Local Similarity   90.8%;  Pred. No. 1.4e-106;
       Matches  413;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |  |  |  |  |  |  ||   ||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047
```

```
RESULT 8
I44908
LOCUS       I44908                    361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
ORIGIN               /mol_type="unassigned DNA"

  Query Match             62.3%;  Score 284.2;  DB 6;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 4.3e-75;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

```
RESULT 9
CSP515700
LOCUS       CSP515700                 2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
```

```
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           61.1%;  Score 278.4;  DB 1;  Length 2889;
  Best Local Similarity  77.2%;  Pred. No. 3e-73;
  Matches  352;  Conservative    0;  Mismatches  101;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||||| |||||||||| |||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||    | || || ||       | | |  | |  |   |      |      ||||||
Db       2179 GG---TTCTGTTCCGCAATGGGGCGGCGGCGGTGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||   | |||||| |||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||    ||||  || | ||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

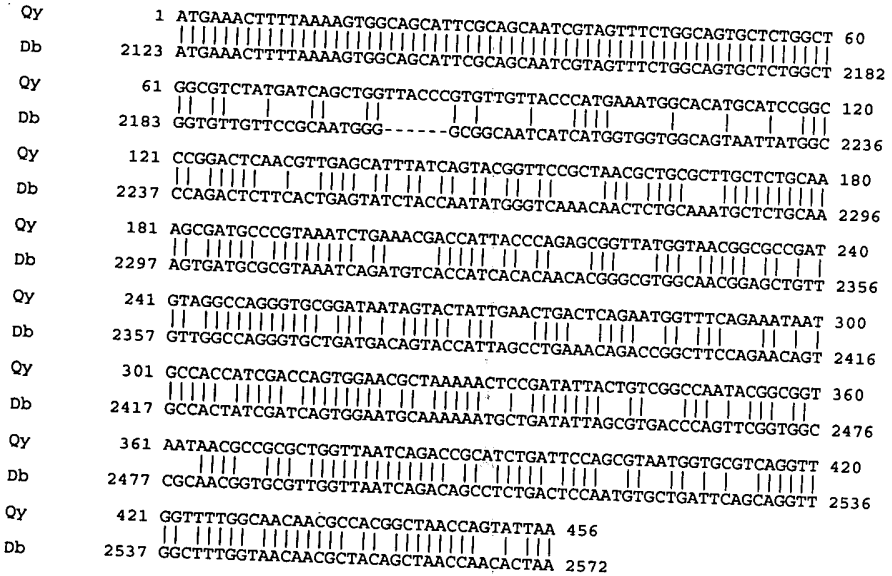
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| ||||||||||  ||||| || || ||||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || ||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571


RESULT 10
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN

  Query Match           50.3%;  Score 229.2;  DB 1;  Length 2920;
  Best Local Similarity  70.6%;  Pred. No. 2.8e-58;
  Matches  322;  Conservative    0;  Mismatches  128;  Indels    6;  Gaps    1;
```



```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
Db       2183 GGTGTTGTTCCGCAATGGG------GCGGCAATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

RESULT 11
AF275733
LOCUS AF275733 1711 bp DNA linear BCT 15-MAY-2001
DEFINITION Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and CsgA protein (csgA) genes, complete cds.
ACCESSION AF275733
VERSION AF275733.1 GI:14039399
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1711)
AUTHORS Uhlich,G.A., Keen,J.E. and Elder,R.O.
TITLE Mutations in the csgD promoter associated with variations in curli expression in certain strains of Escherichia coli O157:H7
JOURNAL Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
MEDLINE 21218556
PUBMED 11319125
REFERENCE 2 (bases 1 to 1711)
AUTHORS Uhlich,G.A., Keen,J.E. and Elder,R.O.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES Location/Qualifiers
source 1. .1711
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="43895 Red Variant"
/serotype="O157:H7"
/db_xref="taxon:562"
gene 757. .1212
/gene="csgB"
CDS 757. .1212
/gene="csgB"
/codon_start=1
/transl_table=11
/product="CsgB protein"
/protein_id="AAK53211.1"
/db_xref="GI:14039400"
/translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
gene 1253. .1711
/gene="csgA"
CDS 1253. .1711
/gene="csgA"
/codon_start=1
/transl_table=11
/product="CsgA protein"
/protein_id="AAK53212.1"
/db_xref="GI:14039401"
/translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
ORIGIN

Query Match 49.4%; Score 225.4; DB 1; Length 1711;
Best Local Similarity 69.7%; Pred. No. 3.8e-57;
Matches 320; Conservative 0; Mismatches 136; Indels 3; Gaps 1;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db       1253 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 1312

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCA---TGAAATGGCACATGCATCC 117
Db       1313 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 1372

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
Db       1373 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 1432

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
Db       1433 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 1492

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
Db       1493 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 1552

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
Db       1553 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 1612

Qy        358 GGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAG 417
Db       1613 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 1672

Qy        418 GTTGGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       1673 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 1711
```

RESULT 12
AE005315
LOCUS AE005315 10190 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139 of 155.
ACCESSION AE005315 AE005174
VERSION AE005315.1 GI:12514572
KEYWORDS .
SOURCE Escherichia coli O157:H7 EDL933
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10190)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,

```
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10190
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            617. .1072
                     /gene="csgB"
                     /note="synonym: Z1675"
     CDS             617. .1072
                     /gene="csgB"
                     /function="structural component; Cell exterior
                     constituents: Surface structures"
                     /note="Residues 1 to 151 of 151 are 100.00 pct identical
                     to residues 1 to 151 of 151 from Escherichia coli K-12
                     Strain MG1655: B1041"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAG55787.1"
                     /db_xref="GI:12514573"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1113. .1571
                     /gene="csgA"
                     /note="synonym: Z1676"
     CDS             1113. .1571
                     /gene="csgA"
                     /function="regulator; Cell exterior constituents: Surface
                     structures"
                     /note="Residues 1 to 152 of 152 are 96.71 pct identical to
                     residues 1 to 151 of 151 from Escherichia coli K-12 Strain
                     MG1655: B1042"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin major subunit, coiled surface structures;
                     cryptic"
                     /protein_id="AAG55788.1"
                     /db_xref="GI:12514574"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
     gene            1630. .1962
                     /gene="csgC"
                     /note="synonym: Z1677"
     CDS             1630. .1962
                     /gene="csgC"
                     /function="putative structure; Not classified"
                     /note="Residues 1 to 110 of 110 are 98.18 pct identical to
                     residues 1 to 110 of 110 from Escherichia coli K-12 Strain
                     MG1655: B1043"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="AAG55789.1"
                     /db_xref="GI:12514575"
                     /translation="MNALLLLAALSSQITFNTTQQGDMYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     gene            2083. .2394
                     /gene="Z1678"
     CDS             2083. .2394
                     /gene="Z1678"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 103 of 103 are 98.05 pct identical to
                     residues 1 to 103 of 103 from Escherichia coli K-12 Strain
                     MG1655: B1044"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55790.1"
                     /db_xref="GI:12514576"
                     /translation="MFRPFLNSLMLGSLFFPFIAIAGSTAQGGVIHFYGQIVEPACDV
                     STQSSPVEMNCPQNGSVPGKTYSSKALMSGNVKNAQIASVKVQYLDKQKKLAVMNIEY
                     N"
     gene            2489. .3022
                     /gene="Z1679"
     CDS             2489. .3022
                     /gene="Z1679"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 177 of 177 are 100.00 pct identical
                     to residues 1 to 177 of 177 from Escherichia coli K-12
                     Strain MG1655: B1045"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyprotein"
                     /protein_id="AAG55791.1"
                     /db_xref="GI:12514577"
                     /translation="MKTRIHVVQGDITKLAVDVIVNAANPSLMGGGGVDGAIHRAAGP
                     ALLDACLKVRQQQGDCPTGHAVITLAGDLPAKAVVHTVGPVWRGGEQNEDQLLQDAYL
                     NSLRLVAANSYTSVAFPAISTGVYGYPRAAAAEIAVKTVSEFITRHALPEQVYFVCYD
                     EENAHLYERLLTQQGDE"
     gene            2964. .4445
                     /gene="ymdC"
                     /note="synonym: Z1680"
     CDS             2964. .4445
                     /gene="ymdC"
                     /function="putative enzyme; Not classified"
                     /note="Residues 1 to 493 of 493 are 99.59 pct identical to
                     residues 1 to 493 of 493 from Escherichia coli K-12 Strain
                     MG1655: B1046"
                     /codon_start=1
                     /transl_table=11
                     /product="putative synthase"
                     /protein_id="AAG55792.1"
                     /db_xref="GI:12514578"
                     /translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPGQCGLFPL
                     EKSLDAFAARYRLAEMSEHTLDVQYYIWQDDMSGRLLFSALLAAAKRGVRVRLLLDDN
                     NTPGLDDILRLLDSHPRIEVRLFNPFSFRLLRPLGYITDFSRLNRRMHNKSFTVDGVV
                     TLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWYCKSVSPLQQVLDV
                     PEGEMADRIELPASWHNDAMTHRYLRKMESSPFINHLVDGTLPLIWAKTRLLSDDPAK
                     GEGKAKRHSLLPQRLFDIMGSPSERIDIISSYFVPTRAGVAQLLRMVRKXVKIAILTN
                     SLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREQSSTLHDRGITGNSGASLHAKT
                     FSIDGKTVFIGSFNFDPRSTLLNTEMGFVIESETLAQLIDKRFIQSQYDAAWQLRLDR
                     WGRINWVDRHAKKEIVLKKEPATSFWKRVMVRLASILPVEWLL"
     gene            complement(4453. .5610)
                     /gene="ymdD"
                     /note="synonym: Z1681"
     CDS             complement(4453. .5610)
                     /gene="ymdD"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 385 of 385 are 99.22 pct identical to
                     residues 1 to 385 of 385 from Escherichia coli K-12 Strain
                     MG1655: B1047"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55793.1"
```

```
                     /db_xref="GI:12514579"
                     /translation="MNPVPAQREYFXDSIRAWLMLLGIPFHISLIYSSHTWHVNSAEP
                     SLWLTLFNDFIHSFRMQVFFVISGYFSYMLFLRYPLKKWWKVRVERVGIPMLTAIPLL
                     TLPQFIMLQYVKGKAESWPGLSLYDKYNTLAWELISHLWFLLVLVVMTTLCVWIFKRI
                     RNNLENSDKTNKKFSMVKLSVIFLCLGIGYAVIRRTIFIVYPPILSNGTFNFIVMQTL
                     FYLPFFILGALAFIFPHLKALFTTPSRGCTLAAALAFVAYLLNQRYGSGDAWMYETES
                     VITMVLGLWMVNVVFSFGHRLLNFQSARVTYFVNASLFIYLVHHPLTLFFGAYITPHI
                     TSNWLGFLCGLIFVVGIAIILYEIHLRIPLLKFLFSGKPVVKRENDKAPAR"
     gene            6003. .7538
                     /gene="mdoG"
                     /note="synonym: Z1683"
     CDS             6003. .7538
                     /gene="mdoG"
                     /function="enzyme; Osmotic adaptation"
                     /note="Residues 1 to 511 of 511 are 99.80 pct identical to
                     residues 1 to 511 of 511 from Escherichia coli K-12 Strain
                     MG1655: B1048"
                     /codon_start=1
                     /transl_table=11
                     /product="periplasmic glucans biosynthesis protein"
                     /protein_id="AAG55794.1"
                     /db_xref="GI:12514580"
                     /translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAQSLAGKGYEAPK
                     SNLPSVFRDMKYADYQQIQFNHDKAYWNNLKTPFKLEFYHQGMYFDTPVKINEVTATA
                     VKRIKYSPDYFTFGDVQHDKDTVKDLGFAGFKVLYPINSKDKNDEIVSMLGASYFRVI
                     GAGQVYGLSARGLAIDTALPSGEEFPRFKEFWIERPKPTDKRLTIYALLDSPRATGAY
                     KFVVMPGRDTVVDVQSKIYLRDKVGKLGVAPLTSMFLFGPNQPSPANNYRPELHDSNG
                     LSIHAGNGEWIWRPLNNPKHLAVSSFSMENPQGFGLLQRGRDFSRFEDLDDRYDLRPS
                     AWVTPKGEWGKGSVELVEIPTNDETNDNIVAYWTPDQLPEPGKEMNFKYTITFSRDED
                     KLHAPDNAWVQQTRRSTGDVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDTPVTAQTS
                     IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDAKKTTEMRAALVNADQTLSETWSYQLP
                     ANE"
     gene            7531. .10074
                     /gene="mdoH"
                     /note="synonym: Z1684"
     CDS             7531. .10074
                     /gene="mdoH"
                     /function="enzyme; Osmotic adaptation"

  Query Match           49.4%;  Score 225.4;  DB 1;  Length 10190;
  Best Local Similarity  69.7%;  Pred. No. 4.5e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       1113 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 1172

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCA---TGAAATGGCACATGCATCC 117
              || ||  | ||||    |    | ||| | || |||    ||      |   ||     |
Db       1173 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 1232

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              |||||| | |||  | |||  ||||||||||||||||   | |||| |||| |||||||||
Db       1233 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 1292

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||   ||||| ||||| |||||    || |||||||||   |||  |||||||| ||
Db       1293 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 1352

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||| || |||||  | ||| | ||  | || || ||||| ||   ||| ||  | ||
Db       1353 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 1412

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              |  |||||  | || |||||||||| |||| ||||  |||| || ||     |||| |||
Db       1413 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 1472

Qy        358 GGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAG 417
              ||   ||||  ||    |||| | ||||| |||||| |||||| ||| |  ||   ||||
Db       1473 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 1532

Qy        418 GTTGGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||| ||||| |||||||| || ||| | ||||| |||

Db       1533 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 1571


RESULT 13
AP002554
LOCUS       AP002554               327773 bp    DNA     linear   BCT 21-DEC-2002
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION   AP002554 BA000007
VERSION     AP002554.1  GI:13360491
KEYWORDS    .
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
   PUBMED   10734605
REFERENCE   2
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
   PUBMED   11108008
REFERENCE   3
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
   PUBMED   11111050
REFERENCE   4
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
   PUBMED   11258796
REFERENCE   5  (bases 1 to 327773)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .327773
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
```

```
                /db_xref="taxon:83334"
gene            150. .2258
                /gene="ECs1032"
CDS             150. .2258
                /gene="ECs1032"
                /note="similar to YCBY_ECOLI gi|1787181 percent identity
                99 in 702 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="putative oxidoreductase"
                /protein_id="BAB34455.1"
                /db_xref="GI:13360492"
                /translation="MNSLPASTARGLEELLKTELENLGAVECQVVQGGVHFKGDTRLV
                YQSLMWSRLASRIMLPLGECKVYSDLDLYLGVQAINWTEMFNPGATFAVHFSGLNDTI
                RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRDAPDIRVNVWLHKETASIALDLSGDGLH
                LRGYRDRAGIAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMLATDRAPGL
                HRGRWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIQRARTNAR
                LAGIGELITFEVKDVAQLTNPLPKEPYGTVLSNPPYGERLDSEPALIALHSLLGRIMK
                NQFGGWNLSLFSASPDLLSCLQLRADKQYKAKNGPLDCVQKNYHVAESTPDSKPAMVA
                EDYANRLRKNLKKPFEKWARQEGIECYRLYDADLPEYNVAVDRYADWVVVQEYAPPKTI
                DAHKARQRLFDIIAATISVLGIAPNKLVLKTRERQKGKNQYQKLGEKGEFLEVTEYNA
                HLWVNLTDYLDTGLFLDHRIARRMLGQMSKGKDFLNLFSYTGSATVHAGLGGARSTTT
                VDMSRTYLEWAERNLRLNGLTGRAHRLIQADCLAWLREANEQFDLIFIDPPTFSNSKR
                MEDAFDVQRDHLALMKDLKRLLRAGGTIMFSNNKRGFRMDLDGLAKLGLKAQEITQKT
                LSQDFARNRQIHNCWLITAA"
gene            2270. .4177
                /gene="ECs1033"
CDS             2270. .4177
                /gene="ECs1033"
                /note="similar to UUP_ECOLI gi|1787182 percent identity 99
                in 635 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="putative ATP-binding component of a transport
                system"
                /protein_id="BAB34456.1"
                /db_xref="GI:13360493"
                /translation="MSLISMHGAWLSFSDAPLLDNAELHIEDNERVCLVGRNGAGKST
                LMKILNREQGLDDGRIIYEQDLIVARLQQDPPRNVEGSVYDFVAEGIEEQAEYLKRYH
                DISRLVMNDPSEKNLNELAKVQEQLDHHNLWQLENRINEVLAQLGLDPNVALSSLSGG
                WLRKAALGRALVSNPRVLLLDEPTNHLDIETIDWLEGFLKTFNGTIIFISHDRSFIRN
                MATRIVDLDRGKLVTYPGNYDQYLLEKEEALRVEELQNAEFDRKLAQEEVWIRQGIKA
                RRTRNEGRVRALKAMRRERGERREVMGTAKMQVEEASRSGKIVFEMEDVCYQVDGKQL
                VKDFSAQVLRGDKIALIGPNGCGKTTLLKLMLGQLQADSGRIHVGTKLEVAYFDQHRA
                ELDPDKTVMDNLAEGKQEVMVNGKPRHVLGYLQDFLFHPKRAMTPVRALSGGERNRLL
                LARLFLKPSNLLILDEPTNDLDVETLELLEELIDSYQGTVLLVSHDRQFVDNTVTECW
                IFEGGGKIGRYVGGYHDARGQQEQYVALKQPAVKKNEEPAAPKAETVKRSSSKLSYKL
                QRELEQLPQLLEDLEAKLEALQTQVADASFFSQPHEQTQKVLADMAAAEQELEQAFER
                WEYLEALKNGG"
gene            4307. .5560
                /gene="ECs1034"
CDS             4307. .5560
                /gene="ECs1034"
                /note="similar to PQIA_ECOLI gi|1787183 percent identity
                99 in 417 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="paraquat-inducible protein A"
                /protein_id="BAB34457.1"
                /db_xref="GI:13360494"
                /translation="MCEHHHAAKHILCSQCDMLVALPRLEHGQKAACPRCGTTLTVAW
                DAPRQRPTAYALAALFMLLLSNLFPFVNMNVAGVTSEITLLEIPGVLFSEDYASLGTF
                FLLFVQLVPAFCLITILLLVNRAELPVRLKEQLARVLFQLKTWGMAEIFLAGVLVSFV
                KLMAYGSIGVGSSFLPWCLFCVLQLRAFQCVDRRWLWDDIAPMPELRQPLKPGVTGIR
                QGLRSCSCCTAILPADEPVCPRCGTKGYVRRRNSLQWTLALLVTSIMLYLPANILPIM
                VTDLLGSKMPSTILAGVILLWSEGSYPVAAVIFLASIMVPTLKMIAIAWLCWDAKGHG
                KRDSERMHLIYEVVEFVGRWSMIDVFVIAVLSALVRMGGLMSIYPAMGALMFALVVIM
                TMFSAMTFDPRLSWDRQPESEHEES"
gene            5565. .7205
                /gene="ECs1035"
CDS             5565. .7205
                /gene="ECs1035"
                /note="similar to PQIB_ECOLI gi|1787184 percent identity
                99 in 546 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="paraquat-inducible protein B"
                /protein_id="BAB34458.1"
                /db_xref="GI:13360495"
                /translation="MESNNGEAKIQKVKNWSPVWIFPIVTALIGAWVLFYHYSHQGPE
                VTLITANAEGIEGGKTTIKSRSVDVGVVESATLADDLTHVEIKARLNSGMEKLLHKDT
                VFWVVKPQIGREGISGLGTLLSGVYIELQPGAKGSKMDKYDLLDSPPLAPPDAKGIRV
                VLDSKKAGQLSPGDPVLFRGYRVGSVETSTFDTQKRNISYQLFINAPYDRLVTSNVRF
                WKDSGIAVDLTSAGMRVEMGSLTTLLSGGVSFDVPEGLDLGQPVAPKTAFVLYDDQKS
                IQDSLYTDHIDYLMFFKDSVRGLQPGAPVEFRGIRLGTVSKVPFFAPNMRQTFNDDYR
                IPVLIRIEPERLKMQLGENADVVEHLGELLKRGLRGSLKTGNLVTGALYVDLDFYPNT
                PAITGIREFNGYQIIPTVSGGLAQIQQRLMEALDKINKLPLNPMIEQATSTLSESQRT
                MKNLQTTLDSMNKILASQSMQQLPTDMQSTLRELNRSMQGFQPGSAAYNKMVADMQRL
                DQVLRELQPVLKTLNEKSNALVFEAKDKKDPEPKRAKQ"
gene            7217. .7765
                /gene="ECs1036"
CDS             7217. .7765
                /gene="ECs1036"
                /note="similar to YMBA_ECOLI gi|1787185 percent identity
                98 in 182 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB34459.1"
                /db_xref="GI:13360496"
                /translation="MTIAALWLAGCSSGEINKNYYQLPVVQSGTQSTASQGNRLLWVE
                QVAVPDYLAGNGVVYQTSDVKYVIANNNLWASPLDQQLRNTLVANLSTQLPGWVVASQ
                PLGSAQDTLNVTVTEFNGRYDGKVIVSGEWLLNHQGQLIKRPFRLEGVQTQDGYDEMV
                KVLAGVWSQEAASIAQEIKRLP"
gene            8021. .8188
                /gene="ECs1037"
CDS             8021. .8188
                /gene="ECs1037"
                /note="identical to RMF_ECOLI gi|1787186 (Conserved in
                E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="ribosome modulation factor"
                /protein_id="BAB34460.1"
                /db_xref="GI:13360497"
                /translation="MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGW
                REAMADRVVMA"
gene            complement(8258. .8776)
                /gene="ECs1038"
CDS             complement(8258. .8776)
                /gene="ECs1038"
                /note="similar to FABA_ECOLI gi|1787187 percent identity
                100 in 172 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11

  Query Match             49.4%;  Score 225.4;  DB 1;  Length 327773;
  Best Local Similarity   69.7%;  Pred. No. 6e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||| | || || || ||||||||
Db     326421 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 326480

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCA---TGAAATGGCACATGCATCC 117
                || ||   | ||||   |    | |||  | || |||   ||     |   ||     |
Db     326481 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 326540

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
```

```
                ||||||  |  |||   |  |||  ||||||||||||||||    | ||||  |||| |||||||||
Db      326541 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 326600

Qy         178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
                |||   |||||  |||||  |||||     ||  |||||||||    |||  |||||||||  ||
Db      326601 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 326660

Qy         238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
                |||||  ||  |||||   |  |||  |  ||  |  ||  ||  |||||  ||    |||  ||   |  ||
Db      326661 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 326720

Qy         298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
                |  |||||   |  ||  ||||||||||  ||||  ||||   ||||  ||  ||    ||||  |||
Db      326721 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 326780

Qy         358 GGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAG 417
                ||    ||||   ||   ||||  |  |||||  ||||||  ||||||  |||  |   ||   ||||
Db      326781 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 326840

Qy         418 GTTGGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
                |||||  |||||  ||||||||  ||  |||  |  |||||  |||
Db      326841 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 326879


RESULT 14
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
```

```
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           49.1%;  Score 224;  DB 1;  Length 4680;
  Best Local Similarity   68.2%;  Pred. No. 1.1e-56;
  Matches  311;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||   |  ||||   |    | | | |     |  ||     |   ||     ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184

RESULT 15
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
```

```
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
     promoter        /note="factor Sigma70; predicted +1 start at 1094689"
                     5. .32
     gene            /note="factor Sigma70; predicted +1 start at 1094716"
                     69. .392
     CDS             /gene="b1028"
                     69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
     promoter        LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     gene            275. .299
                     /gene="b1028"
     CDS             /note="factor Sigma32; predicted +1 start at 1094983"
                     389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
                     389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
     protein_bind    FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
     promoter        /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     gene            1253. .1280
                     /gene="ycdU"
     CDS             /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
                     1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
                     1494. .1745
                     /gene="b1030"
                     1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
     gene            /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
                     1513. .1926
                     /gene="ycdV"
```

```
     CDS             /note="synonym: b1031"
                     1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
     gene            GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
                     complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
     promoter        /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
                     2269. .2296
     promoter        /note="factor Sigma70; predicted +1 start at 1096980"
                     2338. .2365
     promoter        /note="factor Sigma70; predicted +1 start at 1097049"
                     2355. .2383
     gene            /note="factor Sigma70; predicted +1 start at 1097067"
                     2393. .3370
                     /gene="ycdW"
     CDS             /note="synonym: b1033"
                     2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
     promoter        GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
     promoter        /note="factor Sigma70; predicted +1 start at 1097992"
                     3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
                     3396. .3423
```

```
Query Match             49.1%;  Score 224;  DB 1;  Length 10346;
Best Local Similarity   68.2%;  Pred. No. 1.2e-56;
Matches  311;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
                ||||||||||||||||| |||||| |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
               |||| |  |||| |  | | | | |  || |  |  || |  |||| |  |||  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || || ||  | ||| |||||| |||||||||  || |||| |||| |||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
```

```
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

Search completed: March 18, 2004, 15:02:48
Job time : 2312.32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64624	Aac64624 AgfA::PT3
	2	403.2	88.4	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	403.2	88.4	456	2	AAT74142	Aat74142 Salmonell
	4	403.2	88.4	456	3	AAC64617	Aac64617 Salmonell
	5	360	78.9	456	3	AAC64626	Aac64626 AgfA::PT3
	6	353.6	77.5	456	3	AAC64625	Aac64625 AgfA::PT3
	7	352	77.2	456	3	AAC64628	Aac64628 AgfA::PT3
	8	352	77.2	456	3	AAC64622	Aac64622 AgfA::PT3
	9	352	77.2	456	3	AAC64629	Aac64629 AgfA::PT3
	10	352	77.2	456	3	AAC64623	Aac64623 AgfA::PT3
	11	348.8	76.5	456	3	AAC64630	Aac64630 AgfA::PT3
	12	348.8	76.5	456	3	AAC64627	Aac64627 AgfA::PT3
	13	345.6	75.8	456	3	AAC64631	Aac64631 AgfA::PT3
	14	284.2	62.3	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	284.2	62.3	361	2	AAT74141	Aat74141 Salmonell
	16	222.4	48.8	456	3	AAC64619	Aac64619 Escherich
	17	222.4	48.8	456	9	ACF36153	Acf36153 E. coli C
	18	210	46.1	646	2	AAQ62647	Aaq62647 Fibronect
	19	149.6	32.8	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
c	21	49	10.7	78	3	AAC64609	Aac64609 AgfA (SEF
	22	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	23	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	24	48	10.5	78	3	AAC64610	Aac64610 AgfA (SEF
	25	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	26	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
	27	41.6	9.1	72	9	ACF36148	Acf36148 E. coli C
c	28	41.2	9.0	78	3	AAC64605	Aac64605 SefA (SEF
	29	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
	30	40.6	8.9	78	3	AAC64606	Aac64606 SefA (SEF
	31	35.6	7.8	1965	4	AAS52627	Aas52627 E. coli D
	32	35.6	7.8	1965	7	ACA32662	Aca32662 Prokaryot
c	33	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	34	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
	35	34.6	7.6	456	3	AAC64620	Aac64620 Escherich
c	36	34.6	7.6	1506	6	ABQ90743	Abq90743 M. capsul
	37	34	7.5	456	3	AAC64618	Aac64618 Salmonell
c	38	34	7.5	2000	7	ADA71938	Ada71938 Rice gene
	39	34	7.5	2583	7	ACA31722	Aca31722 Prokaryot
	40	34	7.5	2886	7	ACC59886	Acc59886 Saccharom
	41	34	7.5	3300	3	AAZ55699	Aaz55699 DNA encod
	42	33.6	7.4	549	3	AAC95305	Aac95305 Cat flea
	43	33.4	7.3	3990	2	AAX61785	Aax61785 B. burgdo
	44	33.4	7.3	5100	2	AAX20271	Aax20271 Borrelia
	45	33.4	7.3	7766	2	AAT85043	Aat85043 Borrelia

ALIGNMENTS

RESULT 1
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

```
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 8.5e-141;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag=  a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match           88.4%;  Score 403.2;  DB 2;  Length 456;
  Best Local Similarity  92.8%;  Pred. No. 2.9e-123;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |  |    | |    |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
AAT74142
```

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match            88.4%;  Score 403.2;  DB 2;  Length 456;
  Best Local Similarity  92.8%;  Pred. No. 2.9e-123;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |  |    | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

```
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.4%;  Score 403.2;  DB 3;  Length 456;
  Best Local Similarity   92.8%;  Pred. No. 2.9e-123;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match             78.9%;  Score 360;  DB 3;  Length 456;
  Best Local Similarity   86.8%;  Pred. No. 6.4e-109;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||| |||   |   ||||||   | |   | |   ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match           77.5%;  Score 353.6;  DB 3;  Length 456;
  Best Local Similarity  86.0%;  Pred. No. 8.5e-107;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |  ||  ||  | | | |  |  | |  |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||      |  ||  |  | || | |  ||  | | ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.9e-106;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |  ||  ||  | | | |  |  |  |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||     ||  ||    |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||  |  ||  |  |||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 8
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.9e-106;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |  ||  ||  | | | |  |  |  |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             77.2%;  Score 352;  DB 3;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.9e-106;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |.||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.9e-106;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |  |  | |     |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||     | ||||  ||         |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match            76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 3.3e-105;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |   | |     |   ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | |||   |       | |  ||   |    |       |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 3.3e-105;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |   | |     |   ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||  | |       | || |    || | |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             75.8%;  Score 345.6;  DB 3;  Length 456;
  Best Local Similarity   84.9%;  Pred. No. 3.8e-104;
  Matches  387;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   |   | |     |   ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||   || |||  |    |                |    | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             62.3%;  Score 284.2;  DB 2;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 8.4e-84;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             62.3%;  Score 284.2;  DB 2;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 8.4e-84;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 18, 2004, 09:11:57
Job time : 289.405 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	403.2	88.4	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	284.2	62.3	361	1	US-08-233-788A-56	Sequence 56, Appl
	3	33.4	7.3	7766	4	US-09-125-619-3	Sequence 3, Appli
c	4	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	5	31.2	6.8	2436	4	US-09-540-236-492	Sequence 492, App
c	6	31.2	6.8	65792	4	US-09-596-002-31	Sequence 31, Appl
	7	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
	8	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	9	30.2	6.6	1155	4	US-09-328-352-1775	Sequence 1775, Ap
c	10	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
	11	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	12	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, App
c	13	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, App
c	14	29.4	6.4	3494	4	US-09-023-655-1006	Sequence 1006, Ap
c	15	29.2	6.4	2052	1	US-08-443-104-5	Sequence 5, Appli
c	16	29.2	6.4	2052	1	US-08-238-130-6	Sequence 6, Appli
c	17	29.2	6.4	2052	1	US-08-442-859-5	Sequence 5, Appli
c	18	29.2	6.4	2052	2	US-08-398-489-5	Sequence 5, Appli
c	19	29.2	6.4	2052	2	US-08-894-772-1	Sequence 1, Appli
c	20	29.2	6.4	2052	2	US-09-207-844-1	Sequence 1, Appli
c	21	29.2	6.4	2052	4	US-09-252-509-1	Sequence 1, Appli
c	22	29.2	6.4	2052	5	PCT-US95-05534-5	Sequence 5, Appli
c	23	29	6.4	13794	4	US-08-956-171E-54	Sequence 54, Appl
c	24	29	6.4	1830121	4	US-09-557-884-1	Sequence 1, Appli
c	25	29	6.4	1830121	4	US-09-643-990A-1	Sequence 1, Appli
	26	28.8	6.3	720	4	US-09-328-352-354	Sequence 354, App
	27	28.8	6.3	1662	1	US-08-565-386-2	Sequence 2, Appli
	28	28.8	6.3	1836	4	US-09-252-991A-15831	Sequence 15831, A
	29	28.8	6.3	2055	4	US-09-252-991A-15771	Sequence 15771, A
c	30	28.8	6.3	2106	4	US-09-252-991A-15891	Sequence 15891, A
	31	28.8	6.3	4529	1	US-08-565-386-1	Sequence 1, Appli
c	32	28.8	6.3	8906	4	US-09-027-169-5	Sequence 5, Appli
c	33	28.4	6.2	1520	4	US-09-484-970B-30	Sequence 30, Appl
	34	28.2	6.2	2116	3	US-09-000-041A-1	Sequence 1, Appli
	35	28.2	6.2	2426	3	US-09-028-327-2	Sequence 2, Appli
	36	28.2	6.2	2426	4	US-09-571-078A-2	Sequence 2, Appli
	37	28.2	6.2	2812	1	US-08-920-812-16	Sequence 16, Appl
	38	28.2	6.2	2812	1	US-08-920-827-16	Sequence 16, Appl
	39	28.2	6.2	2812	1	US-08-921-177-16	Sequence 16, Appl
	40	28.2	6.2	2812	1	US-08-362-577C-16	Sequence 16, Appl
	41	28.2	6.2	2812	2	US-08-920-828-16	Sequence 16, Appl
c	42	28.2	6.2	4973	3	US-09-381-862-6	Sequence 6, Appli
	43	28.2	6.2	8078	4	US-09-702-251-3	Sequence 3, Appli
	44	28	6.1	906	4	US-09-134-001C-383	Sequence 383, App
c	45	28	6.1	1665	4	US-09-540-236-1858	Sequence 1858, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 88.4%; Score 403.2; DB 1; Length 456;
Best Local Similarity 92.8%; Pred. No. 5e-135;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             62.3%;  Score 284.2;  DB 1;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 2.9e-92;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

```
RESULT 3
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-09-125-619-3

  Query Match              7.3%;  Score 33.4;  DB 4;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 0.33;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |  ||||     ||  ||| |  | | |  | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||     | |   |   | |||||||||  | |||   |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551
```

```
Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |  || ||| |   |   | ||| |           | ||| ||||| |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||       | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 4
5486473-3/c
;Patent No. 5486473
;    APPLICANT: FUKITA, HIROYUKI;YOSHIDA, IWAO;TAKAGI, MITSUO;
;MANABE, SADAO;FUKAI, KONOSUKE
;    TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
;    NUMBER OF SEQUENCES: 16
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,049
;      FILING DATE: 09-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 809,255
;      FILING DATE: 18-DEC-1991
;      APPLICATION NUMBER: 279,685
;      FILING DATE: 05-DEC-1988
;      APPLICATION NUMBER: 932,419
;      FILING DATE: 19-NOV-1986
;SEQ ID NO:3:
;      LENGTH: 1491
5486473-3

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 1491;
  Best Local Similarity   50.0%;  Pred. No. 0.77;
  Matches   78;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||  |  || ||| |  |  | |||    | | | || | | |  | | |  | ||  |
Db        204 CAAGTCACTGACCGAAGCTAGGTAACAGTAACTGCGCACATCTGCGAGGTTGGCTGCTTC 145

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||    ||  ||     ||| ||    |||   |    ||  |||||  || || |
Db        144 CATGTTCATCATTTTGACATCAATGGTTGGCTTGTCTTTTGACATTATGGTCACACAACT 85

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
              |  |||  |    |||||  ||| |    | ||||
Db         84 ATCGCCTTCCAGTACCAGATCAACCCATGTAGCTCC 49


RESULT 5
US-09-540-236-492/c
; Sequence 492, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2005-001
;  CURRENT APPLICATION NUMBER: US/09/540,236
;  CURRENT FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 492
;   LENGTH: 2436
;   TYPE: DNA
;   ORGANISM: M.catarrhalis
US-09-540-236-492

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 2436;
  Best Local Similarity   53.7%;  Pred. No. 1;
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |||||||  |    | ||  |       | | |  | || || ||||| |||   | |||
Db       1511 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 1452

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |      | |  || ||||  |||| |||||| ||||  |||| ||  || |     |
Db       1451 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 1392

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |   || ||  ||  |||| ||  ||||   |  |||
Db       1391 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 1348


RESULT 6
US-09-596-002-31/c
; Sequence 31, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:       Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:       PERL Program
; SEQ ID NO 31
;   LENGTH: 65792
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte template ID No. 6632636 31
;  PUBLICATION INFORMATION:
US-09-596-002-31

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 65792;
  Best Local Similarity   53.7%;  Pred. No. 7.1;
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |||||||  |    | ||  |       | | |  | || || ||||| |||   | |||
Db       6819 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 6760

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |      | |  || ||||  |||| |||||| ||||  |||| ||  || |     |
Db       6759 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 6700

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |   || ||  ||  |||| ||  ||||   |  |||
Db       6699 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 6656


RESULT 7
US-08-508-761B-1
; Sequence 1, Application US/08508761B
; Patent No. 6027920
;  GENERAL INFORMATION:
;    APPLICANT:  Joliff, Gwennael
;    APPLICANT:  Guyonvarch, Armel
;    APPLICANT:  Purification, Relano
;    APPLICANT:  Duchiron, Francis
;    APPLICANT:  Renaud, Michel
;    TITLE OF INVENTION:  System for Protein Expression and
;    TITLE OF INVENTION:  Secretion Especially in Corynebacteria
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jacobson, Price, Holman & Stern, PLLC
;      STREET:  400 Seventh St. N.W.
;      CITY:  Washington D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/508,761B
;      FILING DATE:  31-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09652
;      FILING DATE:  29-JUL-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09870
;      FILING DATE:  02-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Player, William E.
;      REGISTRATION NUMBER:  31,409
;      REFERENCE/DOCKET NUMBER:  P58525NA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 638-6666
;      TELEFAX:  (202) 393-5350
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2547 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Corynebacterium Melassecola
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  420..2390
US-08-508-761B-1

  Query Match              6.8%;  Score 30.8;  DB 3;  Length 2547;
  Best Local Similarity   58.9%;  Pred. No. 1.5;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
              |||| |||||  |  | | |  ||| ||| || ||||   ||||  | ||   ||  |||
Db        876 CAGAGTGGTTGGACCATTGAGACCAACATTGAGCAGTACTACGCCGATAAGAACGCCATT 935

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCC 369
                |||   || |  ||| ||  | | | ||
Db        936 GTTGTGCTCCCAATCGGTGGCGAGAGCTCC 965


RESULT 8
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 6.8%; Score 30.8; DB 4; Length 1664976;
Best Local Similarity 51.4%; Pred. No. 60;
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy        288 TTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGG 347
              ||| | |||| || || || | | |    |||| || ||||    ||||| | ||
Db     360315 TTTTATAAATGTTGTCATGATGGTCAAAATCAACGAAAATAACTGTTTTATTATTCTCAT 360374

Qy        348 CCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAAT 407
              | |    |    || |||  || | | || || | | | | |   ||    ||
Db     360375 CAACGGTGAAAACAAGAACAAAGCTTTTATCGATATGAACTCTTTTAAAATCATTTAAAG 360434

Qy        408 GGTGCGTCAGGTTGGTTT 425
              | ||  | |||||  | |
Db     360435 GATGTCTTAGGTTTTTAT 360452
```

RESULT 9
US-09-328-352-1775
; Sequence 1775, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1775
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1775

Query Match 6.6%; Score 30.2; DB 4; Length 1155;
Best Local Similarity 51.9%; Pred. No. 1.5;
Matches 68; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy         79 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGC 138
              | ||  || |    ||  || .|| ||||  ||| ||  |||   |||  ||| | |
Db        799 GATATTCGAGCAAATAATCAGCAATTGGCTTATGGATATGGCTATGACCAAACTTGGGTA 858

Qy        139 ATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCT 198
              ||| ||||  |   |   | ||       |||||          | || ||  ||||||
Db        859 ATTAATCAAAAGTCTCAAGGGAAACTCAATCTTGCAGCTATTGTGGTTGACCCAAAATCT 918

Qy        199 GAAACGACCAT 209
               ||  ||||||
Db        919 AAACGGACCAT 929
```

RESULT 10
US-09-328-352-2055/c
; Sequence 2055, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2055

```
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-2055

  Query Match             6.5%;  Score 29.6;  DB 4;  Length 972;
  Best Local Similarity   61.8%;  Pred. No. 2.2;
  Matches   47;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || |||| | | | | ||| |   ||  ||||| |||| |   | | |||| | | ||
Db        558 ATTAGTGAAGTAAATAATAATTGCACCAATAATAACGCCAACTGAGCCTAGTGCATCACT 499

Qy        325 AAAAACTCCGATATTA 340
              || ||| | | || |
Db        498 CAATACTTCTAGATAA 483


RESULT 11
US-08-961-527-310
; Sequence 310, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  310:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1939 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-310

  Query Match             6.4%;  Score 29.4;  DB 4;  Length 1939;
  Best Local Similarity   56.8%;  Pred. No. 4;
  Matches   54;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              ||| |||||| |  ||.| |||| ||       || |||| |  || ||| ||    |||
Db        618 GCCAATGTAGTCACGGTTACGGAAAAAGACTTTACTGAATTTCCTAAGAGTGACACCAGT 677

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAA 329
              ||  | | | | | ||     | |  ||| ||||
Db        678 AAGGTTCCTATCCTAGATAGAAGTACTGCTGAAAA 712
```

```
RESULT 12
US-09-139-802-200/c
; Sequence 200, Application US/09139802
; Patent No. 6180084
; GENERAL INFORMATION:
;  APPLICANT: Ruoslahti, Erkki
;  APPLICANT: Pasqualini, Renata
;  TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing
;  TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using
;  TITLE OF INVENTION: Same
;  FILE REFERENCE: P-LJ 3203
;  CURRENT APPLICATION NUMBER: US/09/139,802
;  CURRENT FILING DATE: 1998-08-25
;  EARLIER APPLICATION NUMBER: 08/926,914
;  EARLIER FILING DATE: 1997-09-10
;  EARLIER APPLICATION NUMBER: 08/710,067
;  EARLIER FILING DATE: 1996-09-10
;  NUMBER OF SEQ ID NOS: 226
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 200
;   LENGTH: 3494
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (121)..(3024)
US-09-139-802-200

  Query Match             6.4%;  Score 29.4;  DB 3;  Length 3494;
  Best Local Similarity   45.2%;  Pred. No. 5.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   |  | || || |  |||| || ||   || |   ||||
Db       2531 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 2472

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              |  ||   ||    |    | || ||||    | | |     |        ||||
Db       2471 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 2412

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||      | |||  |||   ||| |  |||||  |||||      |
Db       2411 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 2352

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||     ||  |    |  ||  | |     |  | ||| |
Db       2351 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 2293


RESULT 13
US-09-659-786-200/c
; Sequence 200, Application US/09659786
; Patent No. 6491894
; GENERAL INFORMATION:
;  APPLICANT:  Ruoslahti, Erkki
;  APPLICANT:  Pasqualini, Renata
;  TITLE OF INVENTION:  NGR Receptor and Methods of Identifying Tumor Homing
;  TITLE OF INVENTION:  Molecules That Home to Angiogenic Vasculature Using
;  TITLE OF INVENTION:  Same
;  FILE REFERENCE:  P-LJ 3203
;  CURRENT APPLICATION NUMBER: US/09/659,786
;  CURRENT FILING DATE:  2000-09-11
;  PRIOR APPLICATION NUMBER:  08/926,914
;  PRIOR FILING DATE:  1997-09-10
;  PRIOR APPLICATION NUMBER:  08/710,067
;  PRIOR FILING DATE:  1996-09-10
;  NUMBER OF SEQ ID NOS:  226
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 200
;   LENGTH: 3494
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (121)..(3024)
US-09-659-786-200

  Query Match             6.4%;  Score 29.4;  DB 4;  Length 3494;
  Best Local Similarity   45.2%;  Pred. No. 5.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   | | || || |   |||| || ||   || |   ||||
Db       2531 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 2472

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              |  ||   ||    |     | || ||||    | | |     |         ||||
Db       2471 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 2412

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  |||||  |||||      |
Db       2411 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 2352

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||    ||  |   |  ||  | |    |  | ||| |
Db       2351 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 2293


RESULT 14
US-09-023-655-1006/c
; Sequence 1006, Application US/09023655
; Patent No. 6607879
;  GENERAL INFORMATION:
;    APPLICANT:  Cocks, Benjamin G.
;    APPLICANT:  Susan G. Stuart
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  1508
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,655
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  1006:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3494 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GENBANK
;      CLONE:  g178535
US-09-023-655-1006

  Query Match             6.4%;  Score 29.4;  DB 4;  Length 3494;
  Best Local Similarity   45.2%;  Pred. No. 5.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   | | || || |   |||| || ||   || |   ||||
Db       2531 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 2472

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              |  ||   ||    |     | || ||||    | | |     |         ||||
Db       2471 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 2412

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  |||||  |||||      |
Db       2411 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 2352

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||    ||  |   |  ||  | |    |  | ||| |
Db       2351 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 2293


RESULT 15
US-08-443-104-5/c
; Sequence 5, Application US/08443104
; Patent No. 5691162
;  GENERAL INFORMATION:
;    APPLICANT:  Shuster, Jeffrey R.
;    APPLICANT:  Madden, Mark
;    APPLICANT:  Moyer, Donna L.
;    APPLICANT:  Fuglsang, Claus
;    APPLICANT:  Branner, Sven
;    TITLE OF INVENTION:  NOVEL METALLOPROTEASE HAVING INCREASED
;    TITLE OF INVENTION:  ACTIVITY
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5691162o No. 5691162disk of No. 5691162th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Tape
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/443,104
;      FILING DATE:  17-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/398,489
;      FILING DATE:  03=MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4180.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2052 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-443-104-5

  Query Match             6.4%;  Score 29.2;  DB 1;  Length 2052;
  Best Local Similarity   49.4%;  Pred. No. 4.9;
  Matches   76;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;
```

```
Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGG 359
              || || | || ||  |  ||     || | |  || | |   ||   ||    |||| |
Db        563 TGTCAGCCTCAACGGGGAGAGAAAGAACATCAACGGTGTCCTTGAGAGCGTCGACGGGGT 504

Qy        360 TAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGT 419
                |  || | |||| | |||   |||||    |  | ||  ||||   |    |   ||
Db        503 CTTTCTCGTCACGCTTGGTAAGAGGACCGGGAATCTTGCCCTCGTAGAAGCTGTTTCCGT 444

Qy        420 TGGTTTTGGCAACAACGCCACGGCTAACCAGTAT 453
               ||    | |  |  || ||  || || | ||||
Db        443 AGGAGAAGACCTCGCCGTCAGCGCCAATCTGTAT 410
```

Search completed: March 18, 2004, 20:20:36
Job time : 57.4971 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	35.6	7.8	1965	9	US-09-815-242-6264	Sequence 6264, Ap
	2	35.6	7.8	1965	12	US-10-282-122A-20532	Sequence 20532, A
	3	34	7.5	2583	12	US-10-282-122A-19592	Sequence 19592, A
	4	34	7.5	2886	9	US-09-801-368-131	Sequence 131, App
	5	33.6	7.4	549	10	US-09-991-936-1800	Sequence 1800, Ap
	6	33.4	7.3	7766	12	US-10-222-566-3	Sequence 3, Appli
	7	33.4	7.3	7766	14	US-10-222-162-3	Sequence 3, Appli
	8	33.4	7.3	7766	14	US-10-143-024-3	Sequence 3, Appli
	9	32.8	7.2	1059	9	US-09-738-626-482	Sequence 482, App
c	10	32.8	7.2	1113	9	US-09-738-626-481	Sequence 481, App
c	11	32.8	7.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
	12	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
	13	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c	14	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
c	15	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
	16	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
	17	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
	18	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c	19	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
c	20	31.2	6.8	2394	12	US-10-282-122A-26837	Sequence 26837, A
	21	31.2	6.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	22	31	6.8	525	10	US-09-918-995-32144	Sequence 32144, A
c	23	31	6.8	616	12	US-10-424-599-88945	Sequence 88945, A
	24	31	6.8	2751	15	US-10-398-221-3563	Sequence 3563, Ap
	25	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
c	26	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
	27	30.8	6.8	1665	15	US-10-369-493-43454	Sequence 43454, A
	28	30.8	6.8	2050	14	US-10-086-510-1	Sequence 1, Appli
	29	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appli
	30	30.6	6.7	2301	9	US-09-815-242-4844	Sequence 4844, Ap
	31	30.6	6.7	2301	9	US-09-815-242-9042	Sequence 9042, Ap
c	32	30.4	6.7	392	12	US-10-282-122A-25231	Sequence 25231, A
c	33	30.4	6.7	25220	11	US-09-984-429-364	Sequence 364, App
	34	30.2	6.6	379	12	US-10-424-599-102524	Sequence 102524,
	35	30.2	6.6	3203	12	US-10-282-122A-24039	Sequence 24039, A
	36	30.2	6.6	3222	12	US-10-282-122A-6780	Sequence 6780, Ap
	37	30.2	6.6	3222	14	US-10-210-115-19	Sequence 19, Appl
	38	30.2	6.6	3222	15	US-10-369-493-24379	Sequence 24379, A
c	39	30.2	6.6	1601042	15	US-10-027-632-59064	Sequence 59064, A
c	40	30	6.6	414	14	US-10-156-761-2947	Sequence 2947, Ap
c	41	30	6.6	1074	12	US-10-282-122A-30822	Sequence 30822, A
	42	29.8	6.5	436	12	US-10-424-599-73842	Sequence 73842, A
	43	29.8	6.5	815	12	US-10-424-599-72759	Sequence 72759, A
c	44	29.8	6.5	1394	14	US-10-283-423-13	Sequence 13, Appl
c	45	29.8	6.5	1394	14	US-10-213-821-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-6264
; Sequence 6264, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6264
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Escherichia coli

```
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1965)
US-09-815-242-6264

  Query Match             7.8%;  Score 35.6;  DB 9;  Length 1965;
  Best Local Similarity   60.2%;  Pred. No. 0.27;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATG 100
              || ||| |||| ||| | ||||| |||| || |||| |||| |||| | | ||
Db        350 TTGCTGACAGTTCTCGTCCAGGCGTAGGTGATGAGTTGGTCACCCGAATTGTACCGCTTG 409

Qy        101 AAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGC 138
              | |  |  | |||    | || || | |   | |  ||
Db        410 AGAACGTTCCTGCTCGTGACCTGGCCCCCCTGCTCCGC 447


RESULT 2
US-10-282-122A-20532
; Sequence 20532, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20532
;   LENGTH: 1965
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-282-122A-20532

  Query Match             7.8%;  Score 35.6;  DB 12;  Length 1965;
  Best Local Similarity   60.2%;  Pred. No. 0.27;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATG 100
              || ||| |||| ||| | ||||| |||| || |||| |||| |||| | | ||
Db        350 TTGCTGACAGTTCTCGTCCAGGCGTAGGTGATGAGTTGGTCACCCGAATTGTACCGCTTG 409

Qy        101 AAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGC 138
              | |  |  | |||    | || || | |   | |  ||
Db        410 AGAACGTTCCTGCTCGTGACCTGGCCCCCCTGCTCCGC 447


RESULT 3
US-10-282-122A-19592
; Sequence 19592, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19592
;   LENGTH: 2583
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-10-282-122A-19592

  Query Match             7.5%;  Score 34;  DB 12;  Length 2583;
  Best Local Similarity   48.9%;  Pred. No. 1.1;
  Matches   91;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy         68 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGCCCGGACT 127
              ||| ||| | |  | ||||  | | ||  || |  |||   ||      |  | |||
Db          1 ATGCTCAACCGAATTACCGTTCAGCTCCCGGTGGAGGGGCTTCTGTTCTGGAAACTGACG 60

Qy        128 CAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATG 187
              ||    ||  | |  | || || |||| ||||||  | ||||| ||| |     |
Db         61 GGCCGCGAGGCGATGTCCGAGTCGTTTGCGCTGACGCTGACCGTGCTCGGCACGGACGCG 120

Qy        188 CCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCC 247
              | | | | | |     |  |  |  |||| | ||| ||    |||| | ||  |
Db        121 CGCATCGACCGCAGTAAGCTGCTGGGCCAGCCGGTCACGGTGACCATCCCGACGCAGAGC 180
```

```
Qy        248 AGGGTG 253
              | ||
Db        181 CTGCTG 186


RESULT 4
US-09-801-368-131
; Sequence 131, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 131
;   LENGTH: 2886
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-131

  Query Match             7.5%;  Score 34;  DB 9;  Length 2886;
  Best Local Similarity   50.6%;  Pred. No. 1.2;
  Matches   82;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              | ||  | | | | || | | | |||||    | ||    || |   |    |||| | |
Db       1501 AACAACAACAATAACGATAACGATAACGATAACAATAATAGCAATAATAATAATAACAAT 1560

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || | |    | | | |||  |  ||  ||||||  ||  | | || |    |||  |
Db       1561 AATAATAATAACAATAATAATAATAACAATAATAATAACAATAACAACAACAACAACAAT 1620

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
              || ||     ||| || |       || |   || | |||||
Db       1621 AATAATAATAATAGTAATAATATTAAAAATAACGTTGATAAC 1662


RESULT 5
US-09-991-936-1800
; Sequence 1800, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
;  APPLICANT: Brandt, Kevin S.
;  APPLICANT:  Gaines, Patrick J.
;  APPLICANT:  Stinchcomb, Dan T.
;  APPLICANT:  Wisnewski, Nancy
;  TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
;  FILE REFERENCE: FC-6-C1
;  CURRENT APPLICATION NUMBER: US/09/991,936
;  CURRENT FILING DATE:  2001-11-21
;  PRIOR APPLICATION NUMBER: US/09/543,668
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: 60/128,704
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 1959
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1800
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Ctenocephalides felis
US-09-991-936-1800

  Query Match             7.4%;  Score 33.6;  DB 10;  Length 549;
  Best Local Similarity   61.4%;  Pred. No. 0.67;
  Matches   54;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATG 100
              || ||| |||| |||   | |||||   |||| || |||| |||||  ||||  | | ||
Db        355 TTGCTGACAGTTCTCGTCCAGGCGTAGGTGATGAGTTGGTCACCCGAATTGTACCGCTTG 414

Qy        101 AAATGGCACATGCATCCGGCCCGGACTC 128
              | |  |  | |||    | || |  | |
Db        415 AGAACGTTCCTGCTCGTGACCTGCCCCC 442


RESULT 6
US-10-222-566-3
; Sequence 3, Application US/10222566
; Publication No. US20040044192A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD3
;  CURRENT APPLICATION NUMBER: US/10/222,566
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-566-3

  Query Match             7.3%;  Score 33.4;  DB 12;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||      ||  ||| |  | | | |  | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |   |   | ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |   |  ||| |          | |||  |||||  |      |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||       | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642
```

```
RESULT 7
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD4
;  CURRENT APPLICATION NUMBER: US/10/222,162
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-162-3

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |    ||||      || ||| |  | | |  |  | ||
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||     |  |   | ||||||||| | |||    ||| ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |  |   | ||| |          | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
               || ||||      | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 8
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD1
;  CURRENT APPLICATION NUMBER: US/10/143,024
;  CURRENT FILING DATE:  2002-05-10
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-143-024-3

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |    ||||      || ||| |  | | |  |  | ||
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||     |  |   | ||||||||| | |||    ||| ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |  |   | ||| |          | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
               || ||||      | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 9
US-09-738-626-482
; Sequence 482, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 482
;   LENGTH: 1059
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-482

  Query Match             7.2%;  Score 32.8;  DB 9;  Length 1059;
  Best Local Similarity   47.2%;  Pred. No. 1.8;
  Matches  100;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | |  |    | | |  |||||  ||    |||  ||  | |||||    |||||| |
Db         76 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 135

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |  |      | | |  ||  | |        || ||  ||     |    |||
Db        136 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 195

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
```

```
Db        196 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 255

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              |  | ||  |||  |  ||  |  | ||||
Db        256 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 287


RESULT 10
US-09-738-626-481/c
; Sequence 481, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 481
;   LENGTH: 1113
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-481

  Query Match             7.2%;  Score 32.8;  DB 9;  Length 1113;
  Best Local Similarity   47.2%;  Pred. No. 1.9;
  Matches  100;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | |  |    | | |  |||||  ||    |||  ||  | |||||   ||||| |
Db        972 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 913

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |     |  |  |  ||  |  |     || ||  ||     |      |||
Db        912 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 853

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              | || | |   ||  |  |     ||| | ||||||   || |     ||| ||| | |
Db        852 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 793

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              |  | ||  |||  |  ||  |  | ||||
Db        792 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 761


RESULT 11
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
;   LENGTH: 3309400
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

  Query Match             7.2%;  Score 32.8;  DB 9;  Length 3309400;
  Best Local Similarity   47.2%;  Pred. No. 1.3e+02;
  Matches  100;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | |  |    | | |  |||||  ||    |||  ||  | |||||   ||||| |
Db     451820 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 451761

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |     |  |  |  ||  |  |     || ||  ||     |      |||
Db     451760 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 451701

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              | || | |   ||  |  |     ||| | ||||||   || |     ||| ||| | |
Db     451700 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 451641

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              |  | ||  |||  |  ||  |  | ||||
Db     451640 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 451609


RESULT 12
US-10-333-631-1
; Sequence 1, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer AG
;  TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
;  TITLE OF INVENTION:  ENZYME
;  FILE REFERENCE: LIO121 Foreign Countries
;  CURRENT APPLICATION NUMBER: US/10/333,631
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US 60/221,917
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/280,112
;  PRIOR FILING DATE: 2001-04-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 568
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90)..(90)
;   OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

  Query Match             7.1%;  Score 32.2;  DB 14;  Length 568;
  Best Local Similarity   52.6%;  Pred. No. 2.1;
  Matches   70;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;
```

```
Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCC 214
              | ||| |    |  ||||| |||||   ||   || ||    ||||||  |  |  ||
Db        103 CTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCTGAAATGGTCCGTCACC 162

Qy        215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              ||  |    || || | || | || | ||| ||||||||    |  |  ||   ||
Db        163 CAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTACCTGGATCTGCGGATA 222

Qy        275 TGACTCAGAATGG 287
              |  ||   |||
Db        223 GCCCACATGCTGG 235


RESULT 13
US-10-333-631-4
; Sequence 4, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer AG
;  TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
;  TITLE OF INVENTION:  ENZYME
;  FILE REFERENCE: LIO121 Foreign Countries
;  CURRENT APPLICATION NUMBER: US/10/333,631
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US 60/221,917
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/280,112
;  PRIOR FILING DATE: 2001-04-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-333-631-4

  Query Match             7.1%;  Score 32.2;  DB 14;  Length 972;
  Best Local Similarity   52.6%;  Pred. No. 2.8;
  Matches   70;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCC 214
              | ||| |    |  ||||| |||||   ||   || ||    ||||||  |  |  ||
Db        202 CTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCTGAAATGGTCCGTCACC 261

Qy        215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              ||  |    || || | || | || | ||| ||||||||    |  |  ||   ||
Db        262 CAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTACCTGGACCTGCGGATA 321

Qy        275 TGACTCAGAATGG 287
              |  ||   |||
Db        322 GCCCACATGCTGG 334


RESULT 14
US-10-369-493-33133/c
; Sequence 33133, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33133
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Xylella fastidiosa
US-10-369-493-33133

  Query Match             7.1%;  Score 32.2;  DB 15;  Length 1083;
  Best Local Similarity   59.1%;  Pred. No. 3;
  Matches   55;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTG 421
              ||  | ||| |||||   |  | |||||     ||  || | ||| | ||| ||||
Db        488 ATCTCACCGAGCTGGGATAAAAAACCGCGCACAATCACACCATAACGCTGCTTCAGCCAC 429

Qy        422 GTTTTGGCAACAACGCCACGGCTAACCAGTATT 454
              |||| ||||| || |||   | ||  | |||
Db        428 TTTTTAGCAACCACCCCAGCAGTCACACGCATT 396


RESULT 15
US-10-282-122A-27869/c
; Sequence 27869, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27869
;   LENGTH: 4830
;   TYPE: DNA
;   ORGANISM: Mycobacterium leprae
US-10-282-122A-27869

  Query Match             7.1%;  Score 32.2;  DB 12;  Length 4830;
  Best Local Similarity   51.0%;  Pred. No. 6.7;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        224 ATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGA 283
```

```
           ||| ||     | |||  | | | |   |||     |    | |  ||| |  |    ||
Db      4229 ATGATAGTAAAGACGACTTCGCCGACATTGCCTCCGACCAGAATCGTGACCGCATCACGA 4170

Qy       284 ATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTG 343
           | |  | |  | ||  |||  || |||||||| | | ||   |  || || | ||    |
Db      4169 ACGCCTGCCCACATGCTGCGGCCTTCGACCAGGGCATCGAGCAGTACACCCAGATCGTCG 4110

Qy       344 TCGGCCAATACGGCGGTAATAACGCCGCG 372
           |||| ||| |||  |  |     | || ||
Db      4109 TCGGTCAAGACGATGTCAGCGGCTCCTCG 4081
```

Search completed: March 19, 2004, 03:49:40
Job time : 261.734 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-15	Sequence 15, Appl
2	403.2	88.4	456	6	US-08-233-642A-56	Sequence 56, Appl
3	403.2	88.4	456	23	US-09-543-407-1	Sequence 1, Appli
4	360	78.9	456	23	US-09-543-407-19	Sequence 19, Appl
5	353.6	77.5	456	23	US-09-543-407-17	Sequence 17, Appl
6	352	77.2	456	23	US-09-543-407-11	Sequence 11, Appl
7	352	77.2	456	23	US-09-543-407-13	Sequence 13, Appl
8	352	77.2	456	23	US-09-543-407-23	Sequence 23, Appl
9	352	77.2	456	23	US-09-543-407-25	Sequence 25, Appl
10	348.8	76.5	456	23	US-09-543-407-21	Sequence 21, Appl
11	348.8	76.5	456	23	US-09-543-407-27	Sequence 27, Appl
12	345.6	75.8	456	23	US-09-543-407-29	Sequence 29, Appl
13	284.2	62.3	361	6	US-08-233-642A-54	Sequence 54, Appl
14	222.4	48.8	456	23	US-09-543-407-3	Sequence 3, Appli
15	222.4	48.8	477	17	US-09-252-691-172	Sequence 172, App
16	222.4	48.8	477	17	US-09-252-691C-172	Sequence 172, App
17	222.4	48.8	477	51	US-10-417-886-172	Sequence 172, App
18	222.4	48.8	675	13	US-08-978-878-3	Sequence 3, Appli
19	222.4	48.8	675	31	US-09-741-873B-3	Sequence 3, Appli
20	220	48.2	453	92	US-60-352-946-1	Sequence 1, Appli
21	220	48.2	453	101	US-60-444-371-1	Sequence 1, Appli
22	180	39.5	396	13	US-08-978-878-1	Sequence 1, Appli
23	155.2	34.0	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
C 28	49	10.7	78	23	US-09-543-407-52	Sequence 52, Appl
29	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
30	48	10.5	78	23	US-09-543-407-53	Sequence 53, Appl
31	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
C 32	41.2	9.0	78	23	US-09-543-407-48	Sequence 48, Appl
33	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
34	40.6	8.9	78	23	US-09-543-407-49	Sequence 49, Appl
35	35.8	7.9	561	17	US-09-252-691-171	Sequence 171, App
36	35.8	7.9	561	17	US-09-252-691C-171	Sequence 171, App
37	35.8	7.9	561	51	US-10-417-886-171	Sequence 171, App
38	35.6	7.8	1965	1	PCT-US02-03987-6264	Sequence 6264, Ap
39	35.6	7.8	1965	33	US-09-815-242-6264	Sequence 6264, Ap
40	35.6	7.8	1965	44	US-10-072-851-6264	Sequence 6264, Ap
41	35.6	7.8	1965	48	US-10-282-122A-20532	Sequence 20532, A
C 42	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 24330, A
C 43	35.4	7.8	3412	1	PCT-US01-08631-24948	Sequence 24948, A
44	34.6	7.6	438	31	US-09-739-449-3383	Sequence 3383, Ap
45	34.6	7.6	438	33	US-09-803-110-3383	Sequence 3383, Ap

ALIGNMENTS

```
RESULT 1
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-139;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.4%;  Score 403.2;  DB 6;  Length 456;
  Best Local Similarity   92.8%;  Pred. No. 6.1e-122;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |    | |    |  ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.4%;  Score 403.2;  DB 23;  Length 456;
  Best Local Similarity   92.8%;  Pred. No. 6.1e-122;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |    | |    |  ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             78.9%;  Score 360;  DB 23;  Length 456;
  Best Local Similarity   86.8%;  Pred. No. 1.2e-107;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| |||    |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match            77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity  86.0%;  Pred. No. 1.6e-105;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |    | |      |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||      |    ||     | | ||  | |     || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match            77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity  85.7%;  Pred. No. 5.3e-105;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |    | |      |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||| |           | | || | ||   |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 5.3e-105;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 5.3e-105;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||  |||       ||  ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 5.3e-105;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |    | |      |   ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |       ||  | ||     ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 6.1e-104;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |    | |      |   ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |   |     | ||  |   || |   |  | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 6.1e-104;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |    | |      |   ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |||  |      | | ||  |   |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 75.8%; Score 345.6; DB 23; Length 456;
Best Local Similarity 84.9%; Pred. No. 6.9e-103;
Matches 387; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||     | | | |  |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |   |                |   | | ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 62.3%; Score 284.2; DB 6; Length 361;
Best Local Similarity 99.0%; Pred. No. 1.3e-82;
Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166
```

```
Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-3

  Query Match             48.8%;  Score 222.4;  DB 23;  Length 456;
  Best Local Similarity   68.0%;  Pred. No. 3.9e-62;
  Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||   | ||||   |   | | | |     | ||      |   ||     ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match             48.8%;  Score 222.4;  DB 17;  Length 477;
  Best Local Similarity   68.0%;  Pred. No. 4e-62;
  Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| |||||||||   | |||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || |      ||  | |   | |     | |                    ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  |||||||||| ||||||||   || |||| | |||||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  ||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||    ||| |||| | ||| |  ||     || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  |||| |||  | ||||| || || |  || |  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  |||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


Search completed: March 19, 2004, 03:01:51
Job time : 2652.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		456	100.0	456	6	I44909	I44909 Sequence 58
2		456	100.0	2067	1	SEU43280	U43280 Salmonella
3		448	98.2	5103	1	STAJ2301	AJ002301 Salmonell
4		448	98.2	22411	1	AE008749	AE008749 Salmonell
5		446.4	97.9	254050	1	AL627269	AL627269 Salmonell
6	c	446.4	97.9	301983	1	AE016840	AE016840 Salmonell
7		440.6	96.6	1048	1	STAGFBA	AJ000514 Salmonell
8		331.8	72.8	361	6	I44908	I44908 Sequence 56
9		310.4	68.1	2889	1	CSP515700	AJ515700 Citrobact
10		262.8	57.6	2920	1	CFR515701	AJ515701 Citrobact
11		256	56.1	4680	1	ECCSGABDG	X90754 E.coli csgG
12		256	56.1	10346	1	AE000205	AE000205 Escherich
13		256	56.1	15047	1	D90741	D90741 Escherichia
14		254.4	55.8	456	6	AX814811	AX814811 Sequence
15		254.4	55.8	648	1	ECOCSGAA	L04979 Escherichia
16		247.8	54.3	1711	1	AF275733	AF275733 Escherich
17		247.8	54.3	10190	1	AE005315	AE005315 Escherich
18		247.8	54.3	306358	1	AE016759	AE016759 Escherich
19		247.8	54.3	327773	1	AP002554	AP002554 Escherich
20		229	50.2	230	1	SEU53207	U53207 Salmonella
21		217.6	47.7	2883	1	ESA515702	AJ515702 Enterobac
22		190.4	41.8	10370	1	AE015131	AE015131 Shigella
23		190.4	41.8	292504	1	AE016981	AE016981 Shigella
24	c	167.4	36.7	437	1	AF237726	AF237726 Shigella
25		93.2	20.4	19201	1	D90742	D90742 Escherichia
26		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27		48.2	10.6	78	6	AX814809	AX814809 Sequence
28		44.8	9.8	72	6	AX814798	AX814798 Sequence
29	c	44	9.6	2000	6	AX655393	AX655393 Sequence
30	c	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
31	c	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
32	c	36	7.9	168267	2	AC146009	AC146009 Pan trogl
33		35.6	7.8	301214	1	AE016786	AE016786 Pseudomon
34		34.6	7.6	11160	1	AE008209	AE008209 Agrobacte
35	c	34.6	7.6	11476	1	AE009405	AE009405 Agrobacte
36		34.4	7.5	151828	2	AC146008	AC146008 Pan trogl
37		34.2	7.5	1458	1	AF173043	AF173043 Orientia
38		34	7.5	956	8	AK058943	AK058943 Oryza sat
39	c	34	7.5	1344	6	BD243128	BD243128 Hypersens
40	c	34	7.5	1344	6	BD243666	BD243666 Hypersens
41	c	34	7.5	1344	6	AR403552	AR403552 Sequence
42	c	34	7.5	1344	6	AX287002	AX287002 Sequence
43	c	34	7.5	1344	6	AX343147	AX343147 Sequence
44	c	34	7.5	1344	6	AX370155	AX370155 Sequence
45	c	34	7.5	1344	6	AX536272	AX536272 Sequence

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           100.0%;  Score 456;  DB 6;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-120;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match           100.0%;  Score 456;  DB 1;  Length 2067;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-120;
```

```
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDPEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match             98.2%;  Score 448;  DB 1;  Length 5103;
  Best Local Similarity   98.9%;  Pred. No. 3.9e-118;
  Matches  451;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
     RBS                complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene               complement(2580. .4088)
                        /gene="STM1128"
     CDS                complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNPFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS                complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene               complement(4413. .5099)
                        /gene="STM1129"
     CDS                complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
     RBS                complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene               5599. .6759
                        /gene="STM1130"
     CDS                5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNPKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene               6793. .7497
                        /gene="STM1131"
     RBS                6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           98.2%;  Score 448;  DB 1;  Length 22411;
  Best Local Similarity  98.9%;  Pred. No. 4.3e-118;
  Matches  451;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

/note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature"
gene 3001. .4011
/gene="STY1079"
/note="synonym: pyrD"
CDS 3001. .4011
/gene="STY1079"
/note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q8Z7S9"
/db_xref="SWISS-PROT:Q8Z7S9"
/translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT HI"
misc_feature 3136. .3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHOdehase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
misc_feature 3238. .3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
misc_feature 3874. .3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
gene 4174. .4719
/gene="STY1080"
CDS 4174. .4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepency in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q8Z7S8"
/translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK IMNDRLKPQVHCHSFSLEQAV"
gene complement(4716. .5825)
/gene="STY1081"
CDS complement(4716. .5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Fasta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q8Z7S7"
/db_xref="SPTREMBL:Q8Z7S7"
/translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
misc_feature complement(4737. .4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
gene 5924. .8032
/gene="STY1082"
CDS 5924. .8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q8Z7S6"
/db_xref="SPTREMBL:Q8Z7S6"
/translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT LSPDFARNRQIHNCWLIRAA"
misc_feature 5927. .7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, score 781.30, E-value 3.7e-231"
misc_feature 6506. .6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
gene 8045. .9952
/gene="STY1083"
CDS 8045. .9952
/gene="STY1083"
/note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa overlap"

Query Match 97.9%; Score 446.4; DB 1; Length 254050;
Best Local Similarity 98.7%; Pred. No. 1.5e-117;
Matches 450; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db 88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db 89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db 89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             97.9%;  Score 446.4;  DB 1;  Length 301983;
Best Local Similarity   98.7%;  Pred. No. 1.5e-117;
  Matches  450;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
     gene            KTAVVVQKQSHMAIRVTQR"
                     593. .1048
     CDS             /gene="agfA"
                     593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
     ORIGIN          IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"

  Query Match             96.6%;  Score 440.6;  DB 1;  Length 1048;
  Best Local Similarity   98.0%;  Pred. No. 4.9e-116;
  Matches  446;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||| |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             72.8%;  Score 331.8;  DB 6;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 1.2e-84;
  Matches  333;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match          68.1%;  Score 310.4;  DB 1;  Length 2889;
  Best Local Similarity  81.6%;  Pred. No. 2.1e-78;
  Matches  372;  Conservative    0;  Mismatches   81;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| ||||||||||||||||||    |     ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  ||||||||||||||||  ||   |  |||||| |||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||    ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || |||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               |||||||||| ||||| || ||||| || || |||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
  Query Match             57.6%;  Score 262.8;  DB 1;  Length 2920;
  Best Local Similarity   75.2%;  Pred. No. 1.2e-64;
  Matches  343;  Conservative    0;  Mismatches  107;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||    ||| ||||  ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||     ||||| || ||   |||  ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||||| ||| | ||||| ||    |||| ||||  || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| ||||| |||||||| || |||||  | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||| |||||||||||| || ||||| |||| ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

```
RESULT 11
ECCSGABDG
LOCUS       ECCSGABDG                 4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
```

```
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match             56.1%;  Score 256;  DB 1;  Length 4680;
  Best Local Similarity   72.6%;  Pred. No. 1.1e-62;
  Matches  331;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy         301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
               || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy         361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy         421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
               || ||||| |||||||| || ||| | ||||| |||
Db        4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 12
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
```

and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

```
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423
```

```
Query Match           56.1%;  Score 256;  DB 1;  Length 10346;
Best Local Similarity  72.6%;  Pred. No. 1.2e-62;
Matches  331;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||  .  | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    |||:|| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

RESULT 13
D90741
LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
FEATURES Location/Qualifiers
source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER

```
     CDS             ILNVSPRRLLNFLESRGMAPIAEFADL"
                     2906. .3460
                     /note="ORF_ID:o231#3
                     similar to PIR Accession Number F64035"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein HI1543"
                     /protein_id="BAA35825.1"
                     /db_xref="GI:4062608"
                     /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                     DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
     CDS             ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                     RTMAPLTRDAISAMWDELEEDSEE"
                     3514. .4053
                     /note="unnamed protein product; ORF_ID:o231#4"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35826.1"
                     /db_xref="GI:4062609"
                     /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                     SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
     gene            IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                     LAARSAKTAHREQEIKNKA"
     CDS             complement(4117. .4950)
                     /gene="csgG"
                     complement(4117. .4950)
                     /gene="csgG"
                     /note="ORF_ID:o231#5
                     similar to SwissProt Accession Number P52103"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgG precursor."
                     /protein_id="BAA35827.1"
                     /db_xref="GI:4062610"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
     gene            VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     CDS             complement(4977. .5393)
                     /gene="csgF"
                     complement(4977. .5393)
                     /gene="csgF"
                     /note="ORF_ID:o231#6
                     similar to SwissProt Accession Number P52104"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgF precursor."
                     /protein_id="BAA35828.1"
                     /db_xref="GI:4062611"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
     gene            AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     CDS             complement(5418. .5807)
                     /gene="csgE"
                     complement(5418. .5807)
                     /gene="csgE"
                     /note="ORF_ID:o231#7
                     similar to SwissProt Accession Number P52105"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgE precursor."
                     /protein_id="BAA35829.1"
                     /db_xref="GI:4062612"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
     gene            RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     CDS             complement(5812. .6462)
                     /gene="csgD"
     CDS             complement(5812. .6462)
                     /gene="csgD"
                     /note="ORF_ID:o231#8
                     similar to SwissProt Accession Number P52106"
                     /codon_start=1
                     /transl_table=11
                     /product="Probable csgAB operon transcriptional regulatory
                     protein."
                     /protein_id="BAA35830.1"
                     /db_xref="GI:4062613"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
     gene            HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     CDS             7217. .7672
                     /gene="csgB"
                     7217. .7672

  Query Match           56.1%;  Score 256;  DB 1;  Length 15047;
  Best Local Similarity  72.6%;  Pred. No. 1.2e-62;
  Matches  331;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 14
AX814811
LOCUS       AX814811                 456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1   456
```

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match             55.8%;  Score 254.4;  DB 6;  Length 456;
  Best Local Similarity   72.4%;  Pred. No. 2.8e-62;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match             55.8%;  Score 254.4;  DB 1;  Length 648;
  Best Local Similarity   72.4%;  Pred. No. 2.9e-62;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 18, 2004, 15:02:22
Job time : 2314.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	2	AAQ87467	Aaq87467 AgfA sequ
	2	456	100.0	456	2	AAT74142	Aat74142 Salmonell
	3	456	100.0	456	3	AAC64617	Aac64617 Salmonell
	4	412.8	90.5	456	3	AAC64626	Aac64626 AgfA::PT3
	5	406.4	89.1	456	3	AAC64625	Aac64625 AgfA::PT3
	6	404.8	88.8	456	3	AAC64628	Aac64628 AgfA::PT3
	7	404.8	88.8	456	3	AAC64622	Aac64622 AgfA::PT3
	8	404.8	88.8	456	3	AAC64629	Aac64629 AgfA::PT3
	9	404.8	88.8	456	3	AAC64623	Aac64623 AgfA::PT3
	10	403.2	88.4	456	3	AAC64624	Aac64624 AgfA::PT3
	11	401.6	88.1	456	3	AAC64630	Aac64630 AgfA::PT3
	12	401.6	88.1	456	3	AAC64627	Aac64627 AgfA::PT3
	13	398.4	87.4	456	3	AAC64631	Aac64631 AgfA::PT3
	14	331.8	72.8	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	331.8	72.8	361	2	AAT74141	Aat74141 Salmonell
	16	254.4	55.8	456	3	AAC64619	Aac64619 Escherich
	17	254.4	55.8	456	9	ACF36153	Acf36153 E. coli C
	18	242	53.1	646	2	AAQ62647	Aaq62647 Fibronect
	19	180	39.5	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
	21	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	22	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
c	23	44	9.6	2000	7	ADA71938	Ada71938 Rice gene
	24	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
	25	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
c	26	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	27	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
	28	34.6	7.6	456	3	AAC64620	Aac64620 Escherich
	29	34	7.5	456	3	AAC64618	Aac64618 Salmonell
c	30	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	31	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	32	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	33	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	34	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	35	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c	36	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a
	37	34	7.5	2886	7	ACC59886	Acc59886 Saccharom
	38	34	7.5	3300	3	AAZ55699	Aaz55699 DNA encod
	39	33.4	7.3	3990	2	AAX61785	Aax61785 B. burgdo
	40	33.4	7.3	5100	2	AAX20271	Aax20271 Borrelia
	41	33.4	7.3	7766	2	AAT85043	Aat85043 Borrelia
	42	33.2	7.3	788	4	AAH08100	Aah08100 Human cDN
c	43	32.8	7.2	78	3	AAC64609	Aac64609 AgfA (SEF
	44	32.8	7.2	1005	7	ACA02133	Aca02133 C. glutam
	45	32.8	7.2	1059	5	AAH65447	Aah65447 C glutami

ALIGNMENTS

RESULT 1
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX
FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)

```
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 456;  DB 2;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 8.2e-138;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
AAT74142
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 456;  DB 2;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 8.2e-138;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
```

PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36341.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.2e-138;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 4
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 3; Length 456;
Best Local Similarity 94.1%; Pred. No. 9.2e-124;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| ||| |  ||||||  | |  | |  || ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 3; Length 456;
Best Local Similarity 93.2%; Pred. No. 1.1e-121;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||         |   ||   |   | ||  | |   || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 3;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 3.7e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||   ||  ||   |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||  |   ||  |  |||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
Best Local Similarity 93.0%; Pred. No. 3.7e-121;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||| |  |||| | || | |  | | | |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |  ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 8
AAC64629
ID AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
Best Local Similarity 93.0%; Pred. No. 3.7e-121;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 9
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

```
  Query Match             88.8%;  Score 404.8;  DB 3;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 3.7e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 88.4%; Score 403.2; DB 3; Length 456;
Best Local Similarity 92.8%; Pred. No. 1.2e-120;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match            88.1%;  Score 401.6;  DB 3;  Length 456;
  Best Local Similarity  92.5%;  Pred. No. 4.1e-120;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |||   |     |  |  ||   |    |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            88.1%;  Score 401.6;  DB 3;  Length 456;
  Best Local Similarity  92.5%;  Pred. No. 4.1e-120;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |  |       | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match           87.4%;  Score 398.4;  DB 3;  Length 456;
  Best Local Similarity   92.1%;  Pred. No. 4.5e-119;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |   |                |    | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             72.8%;  Score 331.8;  DB 2;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-97;
  Matches  333;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             72.8%;  Score 331.8;  DB 2;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-97;
  Matches  333;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 18, 2004, 09:11:45
```

Job time : 291.405 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	331.8	72.8	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	33.4	7.3	7766	4	US-09-125-619-3	Sequence 3, Appli
c	6	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	7	31.2	6.8	2436	4	US-09-540-236-492	Sequence 492, App
c	8	31.2	6.8	65792	4	US-09-596-002-31	Sequence 31, Appl
	9	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
	10	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	11	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	12	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	13	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	14	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	15	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
c	16	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
c	17	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	18	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	19	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	20	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	21	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	22	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, App
c	23	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, App
c	24	29.4	6.4	3494	4	US-09-023-655-1006	Sequence 1006, Ap
c	25	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
	26	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
c	27	29.2	6.4	2052	1	US-08-443-104-5	Sequence 5, Appli
c	28	29.2	6.4	2052	1	US-08-238-130-6	Sequence 6, Appli
c	29	29.2	6.4	2052	1	US-08-442-859-5	Sequence 5, Appli
c	30	29.2	6.4	2052	2	US-08-398-489-5	Sequence 5, Appli
c	31	29.2	6.4	2052	2	US-08-894-772-1	Sequence 1, Appli
c	32	29.2	6.4	2052	2	US-09-207-844-1	Sequence 1, Appli
c	33	29.2	6.4	2052	4	US-09-252-509-1	Sequence 1, Appli
c	34	29.2	6.4	2052	5	PCT-US95-05534-5	Sequence 5, Appli
	35	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	36	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	37	29	6.4	534	4	US-09-489-039A-2195	Sequence 2195, Ap
	38	29	6.4	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
c	39	29	6.4	13794	4	US-08-956-171E-54	Sequence 54, Appl
c	40	28.8	6.3	654	4	US-09-252-991A-4506	Sequence 4506, Ap
	41	28.8	6.3	720	4	US-09-328-352-354	Sequence 354, App
	42	28.8	6.3	1662	1	US-08-565-386-2	Sequence 2, Appli
	43	28.8	6.3	4529	1	US-08-565-386-1	Sequence 1, Appli
c	44	28.8	6.3	8906	4	US-09-027-169-5	Sequence 5, Appli
	45	28.4	6.2	1026	3	US-07-751-891B-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 100.0%; Score 456; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 3e-149;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 72.8%; Score 331.8; DB 1; Length 361;
Best Local Similarity 99.4%; Pred. No. 6e-106;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA AND ITS USE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,927
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,108
; FILING DATE: 06-AUG-1977
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1581
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304

```
;       TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1344 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.12;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.12;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-09-125-619-3

  Query Match             7.3%;  Score 33.4;  DB 4;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 0.51;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |    ||||     ||  ||| |  | | |  | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |   |   | |||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    |  ||| |          | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||        | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 6
5486473-3/c
;Patent No. 5486473
;    APPLICANT: FUKITA, HIROYUKI;YOSHIDA, IWAO;TAKAGI, MITSUO;
;MANABE, SADAO;FUKAI, KONOSUKE
;    TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
;    NUMBER OF SEQUENCES: 16
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,049
;      FILING DATE: 09-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 809,255
;      FILING DATE: 18-DEC-1991
;      APPLICATION NUMBER: 279,685
;      FILING DATE: 05-DEC-1988
;      APPLICATION NUMBER: 932,419
;      FILING DATE: 19-NOV-1986
;SEQ ID NO:3:
;      LENGTH: 1491
5486473-3

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 1491;
  Best Local Similarity   50.0%;  Pred. No. 1.2;
  Matches   78;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||  |  || ||| |  |  | |||    | | | || | | | | | | | | || |
Db        204 CAAGTCACTGACCGAAGCTAGGTAACAGTAACTGCGCACATCTGCGAGGTTGGCTGCTTC 145

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||    ||   ||    ||| ||    |||    |    ||  |||||  || || |
Db        144 CATGTTCATCATTTTGACATCAATGGTTGGCTTGTCTTTTGACATTATGGTCACACAACT 85
```

```
Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
              |  |||  |    |||||   |||  |    | ||||
Db         84 ATCGCCTTCCAGTACCAGATCAACCCATGTAGCTCC 49
```

RESULT 7
US-09-540-236-492/c
; Sequence 492, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 492
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-492

Query Match 6.8%; Score 31.2; DB 4; Length 2436;
Best Local Similarity 53.7%; Pred. No. 1.6;
Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

```
Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |||||||  |    | ||  |      | | |  | || || |||||  |||    | |||
Db       1511 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 1452

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |      | |  || ||||  |||| ||||||  ||||  |||| ||  || |  |
Db       1451 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 1392

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |   || ||  ||  |||| ||  ||||    |  |||
Db       1391 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 1348
```

RESULT 8
US-09-596-002-31/c
; Sequence 31, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 65792
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 31
; PUBLICATION INFORMATION:
US-09-596-002-31

Query Match 6.8%; Score 31.2; DB 4; Length 65792;
Best Local Similarity 53.7%; Pred. No. 9.7;
Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

```
Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |||||||  |    | ||  |      | | |  | || || |||||  |||    | |||
```

```
Db       6819 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 6760

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |      | |  || ||||  |||| ||||||  ||||  |||| ||  || |  |
Db       6759 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 6700

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |   || ||  ||  |||| ||  ||||    |  |||
Db       6699 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 6656
```

RESULT 9
US-08-508-761B-1
; Sequence 1, Application US/08508761B
; Patent No. 6027920
; GENERAL INFORMATION:
; APPLICANT: Joliff, Gwennael
; APPLICANT: Guyonvarch, Armel
; APPLICANT: Purification, Relano
; APPLICANT: Duchiron, Francis
; APPLICANT: Renaud, Michel
; TITLE OF INVENTION: System for Protein Expression and
; TITLE OF INVENTION: Secretion Especially in Corynebacteria
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
; STREET: 400 Seventh St. N.W.
; CITY: Washington D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,761B
; FILING DATE: 31-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/09652
; FILING DATE: 29-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/09870
; FILING DATE: 02-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Player, William E.
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P58525NA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 393-5350
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2547 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium Melassecola
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 420..2390
US-08-508-761B-1

Query Match 6.8%; Score 30.8; DB 3; Length 2547;
Best Local Similarity 58.9%; Pred. No. 2.2;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
              |||| |||||  |  | | |  ||| ||| || ||||   ||||  | ||   ||  |||
```

```
Db        876 CAGAGTGGTTGGACCATTGAGACCAACATTGAGCAGTACTACGCCGATAAGAACGCCATT 935

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |||   || |  ||| ||  | | | ||
Db        936 GTTGTGCTCCCAATCGGTGGCGAGAGCTCC 965


RESULT 10
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             6.8%;  Score 30.8;  DB 4;  Length 1664976;
  Best Local Similarity   51.4%;  Pred. No. 71;
  Matches   71;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        288 TTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGG 347
              ||| | ||||  || ||  || | | |    ||||  || ||||    |||| | ||
Db     360315 TTTTATAAATGTTGTCATGATGGTCAAAATCAACGAAAATAACTGTTTTATTATTCTCAT 360374

Qy        348 CCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAAT 407
              | |    |    || |||   || |  | || || | | | |   ||    ||
Db     360375 CAACGGTGAAAACAAGAACAAAGCTTTTATCGATATGAACTCTTTTAAAATCATTTAAAG 360434

Qy        408 GGTGCGTCAGGTTGGTTT 425
              | ||  | |||||  | |
Db     360435 GATGTCTTAGGTTTTTAT 360452


RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4403765;
  Best Local Similarity   51.9%;  Pred. No. 90;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || |||||||| || ||  |  ||   | ||||| |||| |  |   || ||
Db    1357218 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357159

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
                ||    | | |    ||  | ||| | ||   |   ||  ||  || | |  |||
Db    1357158 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357099

Qy        165 TGCGCTTGCTCTG 177
                ||   |  |||
Db    1357098 CCCGGACGTGCTG 1357086


RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4411529;
  Best Local Similarity   51.9%;  Pred. No. 90;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || |||||||| || ||  |  ||   | ||||| |||| |  |   || ||
Db    1357750 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357691

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
                ||    | | |    ||  | ||| | ||   |   ||  ||  || | |  |||
Db    1357690 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357631

Qy        165 TGCGCTTGCTCTG 177
                ||   |  |||
Db    1357630 CCCGGACGTGCTG 1357618


RESULT 13
US-09-252-991A-10904
; Sequence 10904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
```

```
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10904
;   LENGTH: 1008
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904

  Query Match             6.5%;  Score 29.8;  DB 4;  Length 1008;
  Best Local Similarity   54.0%;  Pred. No. 3;
  Matches   61;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
              | | || | |  |  || | |  | ||| | |   ||||  ||||||  | |     |
Db        141 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 200

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              ||||  |   | ||   || ||||| | |  ||||| || |     || ||| |
Db        201 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 253


RESULT 14
US-09-252-991A-10818/c
; Sequence 10818, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10818
;   LENGTH: 1077
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10818

  Query Match             6.5%;  Score 29.8;  DB 4;  Length 1077;
  Best Local Similarity   54.0%;  Pred. No. 3.1;
  Matches   61;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
              | | || | |  |  || | |  | ||| | |   ||||  ||||||  | |     |
Db        530 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 471

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              ||||  |   | ||   || ||||| | |  ||||| || |     || ||| |
Db        470 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 418


RESULT 15
US-09-252-991A-10865
; Sequence 10865, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10865
;   LENGTH: 1521
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10865

  Query Match             6.5%;  Score 29.8;  DB 4;  Length 1521;
  Best Local Similarity   54.0%;  Pred. No. 3.7;
  Matches   61;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
              | | || | |  |  || | |  | ||| | |   ||||  ||||||  | |     |
Db        170 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 229

Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              ||||  |   | ||   || ||||| | |  ||||| || |     || ||| |
Db        230 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 282


Search completed: March 18, 2004, 20:19:26
Job time : 71.4971 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	34.2	7.5	616	12	US-10-424-599-88945	Sequence 88945, A
c	2	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c	3	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c	4	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c	5	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c	6	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
	7	34	7.5	2886	9	US-09-801-368-131	Sequence 131, App
	8	33.4	7.3	7766	12	US-10-222-566-3	Sequence 3, Appli
	9	33.4	7.3	7766	14	US-10-222-162-3	Sequence 3, Appli
	10	33.4	7.3	7766	14	US-10-143-024-3	Sequence 3, Appli
c	11	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
	12	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	13	32.8	7.2	1059	9	US-09-738-626-482	Sequence 482, App
c	14	32.8	7.2	1113	9	US-09-738-626-481	Sequence 481, App
c	15	32.8	7.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
	16	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
	17	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c	18	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
c	19	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
	20	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
	21	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
	22	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c	23	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
	24	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
	25	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
c	26	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
c	27	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
c	28	31.2	6.8	2394	12	US-10-282-122A-26837	Sequence 26837, A
c	29	31	6.8	525	10	US-09-918-995-32144	Sequence 32144, A
	30	31	6.8	2751	15	US-10-398-221-3563	Sequence 3563, Ap
	31	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
c	32	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
	33	30.8	6.8	2050	14	US-10-086-510-1	Sequence 1, Appli
	34	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
	35	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appli
	36	30.6	6.7	2301	9	US-09-815-242-4844	Sequence 4844, Ap
	37	30.6	6.7	2301	9	US-09-815-242-9042	Sequence 9042, Ap
c	38	30.6	6.7	75216	14	US-10-080-170-646	Sequence 646, App
c	39	30.4	6.7	392	12	US-10-282-122A-25231	Sequence 25231, A
c	40	30.4	6.7	992	12	US-10-425-114-31747	Sequence 31747, A
c	41	30.4	6.7	25220	11	US-09-984-429-364	Sequence 364, App
	42	30.2	6.6	379	12	US-10-424-599-102524	Sequence 102524,
	43	30.2	6.6	654	12	US-10-425-114-28546	Sequence 28546, A
	44	30.2	6.6	3203	12	US-10-282-122A-24039	Sequence 24039, A
	45	30.2	6.6	3222	12	US-10-282-122A-6780	Sequence 6780, Ap

ALIGNMENTS

RESULT 1
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(616)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

Query Match 7.5%; Score 34.2; DB 12; Length 616;
Best Local Similarity 50.9%; Pred. No. 0.46;
Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
              | | |  |||  ||  || ||| || || |||  | |    | |  | | | | |  | |
Db        556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497

Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              | | ||    || |  | |    || ||  |  |    | | |||    ||  |  | |
Db        496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437
```

```
Qy         221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
               || |||||| |  |  ||| |||  ||||| ||| ||
Db         436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTA 398
```

RESULT 2
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.82;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy          52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
               |||| || || ||   | ||  | || |||| || || |  |||| |  || ||||
Db         209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150
Qy         112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
               | ||  | |     ||||| | |  ||  |    |||  | | ||||| |   ||
Db         149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90
Qy         172 GC 173
               ||
Db          89 GC 88
```

RESULT 3
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.82;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy          52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
               |||| || || ||   | ||  | || |||| || || |  |||| |  || ||||
Db         209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150
Qy         112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
               | ||  | |     ||||| | |  ||  |    |||  | | ||||| |   ||
Db         149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90
Qy         172 GC 173
               ||
Db          89 GC 88
```

RESULT 4
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.82;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy          52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
               |||| || || ||   | ||  | || |||| || || |  |||| |  || ||||
Db         209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150
Qy         112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
               | ||  | |     ||||| | |  ||  |    |||  | | ||||| |   ||
Db         149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90
Qy         172 GC 173
               ||
Db          89 GC 88
```

RESULT 5
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora

```
US-10-010-390-6

  Query Match             7.5%;  Score 34;  DB 14;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.82;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 6
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION:  RESISTANCE
;  FILE REFERENCE: 21829/203 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/10/441,736
;  CURRENT FILING DATE:  2003-05-20
;  PRIOR APPLICATION NUMBER: 60/107,243
;  PRIOR FILING DATE: 1998-11-05
;  PRIOR APPLICATION NUMBER: 09/431,614
;  PRIOR FILING DATE: 1999-11-02
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-441-736-5

  Query Match             7.5%;  Score 34;  DB 15;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.82;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 7
US-09-801-368-131
; Sequence 131, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 131
;   LENGTH: 2886
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-131

  Query Match             7.5%;  Score 34;  DB 9;  Length 2886;
  Best Local Similarity   50.6%;  Pred. No. 1.2;
  Matches   82;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              | ||  | | | | || | | |||||    | ||    || |   |    |||| | |
Db       1501 AACAACAACAATAACGATAACGATAACGATAACAATAATAGCAATAATAATAATAACAAT 1560

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || | |    | | | |||  |  ||  ||||||  ||  | | || |    |||  |
Db       1561 AATAATAATAACAATAATAATAATAACAATAATAATAACAATAACAACAACAACAACAAT 1620

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
              || ||     ||| || |        || |   || | |||||
Db       1621 AATAATAATAATAGTAATAATATTAAAAATAACGTTGATAAC 1662


RESULT 8
US-10-222-566-3
; Sequence 3, Application US/10222566
; Publication No. US20040044192A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD3
;  CURRENT APPLICATION NUMBER: US/10/222,566
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-566-3

  Query Match             7.3%;  Score 33.4;  DB 12;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||     ||  ||| |  | | |  |  | |
```

```
Db        4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491
Qy         217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db        4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551
Qy         277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db        4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611
Qy         337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
Db        4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642
```

RESULT 9
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234USD4
; CURRENT APPLICATION NUMBER: US/10/222,162
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/125,619
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-222-162-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy         157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db        4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491
Qy         217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db        4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551
Qy         277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db        4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611
Qy         337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
Db        4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642
```

RESULT 10
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234USD1
; CURRENT APPLICATION NUMBER: US/10/143,024
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 09/125,619
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-143-024-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy         157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db        4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491
Qy         217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db        4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551
Qy         277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db        4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611
Qy         337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
Db        4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642
```

RESULT 11
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1662)
US-10-156-761-3483

Query Match 7.3%; Score 33.2; DB 14; Length 1662;

Best Local Similarity 52.1%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531

Qy         110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

Qy         170 TTGCTCTGCAAAGCGATGCCCG 191
               |||||||||||||||||||||
Db         470 TTGAACTTCTTGGCGGCGACGG 449
```

RESULT 12
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 7.3%; Score 33.2; DB 14; Length 9025608;
Best Local Similarity 52.1%; Pred. No. 1.3e+02;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

Qy         110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

Qy         170 TTGCTCTGCAAAGCGATGCCCG 191
               |||||||||||||||||||||
Db     4331901 TTGAACTTCTTGGCGGCGACGG 4331922
```

RESULT 13
US-09-738-626-482
; Sequence 482, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 482
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-482

Query Match 7.2%; Score 32.8; DB 9; Length 1059;
Best Local Similarity 47.2%; Pred. No. 1.9;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy         184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          76 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 135

Qy         244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         136 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 195

Qy         304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         196 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 255

Qy         364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
               |||||||||||||||||||||||||||||||
Db         256 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 287
```

RESULT 14
US-09-738-626-481/c
; Sequence 481, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 481

```
;   LENGTH: 1113
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-481

  Query Match             7.2%;  Score 32.8;  DB 9;  Length 1113;
  Best Local Similarity   47.2%;  Pred. No. 1.9;
  Matches  100;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | | |   | | |  |||||  ||    |||  ||  | |||||   ||||||
Db        972 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 913

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |    |  | |  ||  | |     || ||  ||     |     |||
Db        912 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 853

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              | || | |   ||  | |     ||| | ||||||   || |   ||| ||| | |
Db        852 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 793

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              |  | ||  |||  |  ||  |  | | ||||
Db        792 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 761


RESULT 15
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
;   LENGTH: 3309400
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

  Query Match             7.2%;  Score 32.8;  DB 9;  Length 3309400;
  Best Local Similarity   47.2%;  Pred. No. 1.2e+02;
  Matches  100;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | | |   | | |  |||||  ||    |||  ||  | |||||   ||||||
Db     451820 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 451761

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |    |  | |  ||  | |     || ||  ||     |     |||
Db     451760 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 451701

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              | || | |   ||  | |     ||| | ||||||   || |   ||| ||| | |
Db     451700 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 451641

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              |  | ||  |||  |  ||  |  | | ||||
Db     451640 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 451609


Search completed: March 19, 2004, 03:48:25
Job time : 272.734 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	6	US-08-233-642A-56	Sequence 56, Appl
2	456	100.0	456	23	US-09-543-407-1	Sequence 1, Appli
3	412.8	90.5	456	23	US-09-543-407-19	Sequence 19, Appl
4	406.4	89.1	456	23	US-09-543-407-17	Sequence 17, Appl
5	404.8	88.8	456	23	US-09-543-407-11	Sequence 11, Appl
6	404.8	88.8	456	23	US-09-543-407-13	Sequence 13, Appl
7	404.8	88.8	456	23	US-09-543-407-23	Sequence 23, Appl
8	404.8	88.8	456	23	US-09-543-407-25	Sequence 25, Appl
9	403.2	88.4	456	23	US-09-543-407-15	Sequence 15, Appl
10	401.6	88.1	456	23	US-09-543-407-21	Sequence 21, Appl
11	401.6	88.1	456	23	US-09-543-407-27	Sequence 27, Appl
12	398.4	87.4	456	23	US-09-543-407-29	Sequence 29, Appl
13	331.8	72.8	361	6	US-08-233-642A-54	Sequence 54, Appl
14	254.4	55.8	456	23	US-09-543-407-3	Sequence 3, Appli
15	254.4	55.8	477	17	US-09-252-691-172	Sequence 172, App
16	254.4	55.8	477	17	US-09-252-691C-172	Sequence 172, App
17	254.4	55.8	477	51	US-10-417-886-172	Sequence 172, App
18	254.4	55.8	675	13	US-08-978-878-3	Sequence 3, Appli
19	254.4	55.8	675	31	US-09-741-873B-3	Sequence 3, Appli
20	252	55.3	453	92	US-60-352-946-1	Sequence 1, Appli
21	252	55.3	453	101	US-60-444-371-1	Sequence 1, Appli
22	210.4	46.1	396	13	US-08-978-878-1	Sequence 1, Appli
23	185.6	40.7	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
c 28	44	9.6	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
29	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
30	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
31	37.4	8.2	1304	51	US-10-437-963-19773	Sequence 19773, A
32	35.8	7.9	561	17	US-09-252-691-171	Sequence 171, App
33	35.8	7.9	561	17	US-09-252-691C-171	Sequence 171, App
34	35.8	7.9	561	51	US-10-417-886-171	Sequence 171, App
c 35	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 24330, A
c 36	35.4	7.8	3412	1	PCT-US01-08631-24948	Sequence 24948, A
37	34.6	7.6	438	31	US-09-739-449-3383	Sequence 3383, Ap
38	34.6	7.6	438	33	US-09-803-110-3383	Sequence 3383, Ap
39	34.6	7.6	456	23	US-09-543-407-4	Sequence 4, Appli
c 40	34.6	7.6	29058	21	US-09-514-000-342	Sequence 342, Ap
41	34.6	7.6	187892	31	US-09-739-449-208	Sequence 208, App
42	34.6	7.6	187892	33	US-09-803-110-208	Sequence 208, App
c 43	34.2	7.5	385	18	US-09-333-535A-288	Sequence 288, App
c 44	34.2	7.5	385	25	US-09-615-606A-909	Sequence 909, App
c 45	34.2	7.5	396	64	US-60-089-524-4522	Sequence 4522, Ap

ALIGNMENTS

```
RESULT 1
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-642A-56

  Query Match           100.0%;  Score 456;  DB 6;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-135;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
```

```
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-135;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             90.5%;  Score 412.8;  DB 23;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 2.2e-121;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||| ||| |  |||||  | | |  | |  || ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             89.1%;  Score 406.4;  DB 23;  Length 456;
  Best Local Similarity   93.2%;  Pred. No. 2.5e-119;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||        |    ||   |      | | ||  | |     || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||| |           | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||          |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 8.3e-119;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||  |||       ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 8.3e-119;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match           88.4%;  Score 403.2;  DB 23;  Length 456;
  Best Local Similarity  92.8%;  Pred. No. 2.7e-118;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |  ||  ||   | | | |  | |   |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match           88.1%;  Score 401.6;  DB 23;  Length 456;
  Best Local Similarity  92.5%;  Pred. No. 8.8e-118;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |  |     | || |   || |  |   | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||          ||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match           88.1%;  Score 401.6;  DB 23;  Length 456;
  Best Local Similarity  92.5%;  Pred. No. 8.8e-118;
  Matches  422;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |||  |     |  |  ||  |    |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 87.4%; Score 398.4; DB 23; Length 456;
Best Local Similarity 92.1%; Pred. No. 9.4e-117;
Matches 420; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |  |                    |    | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 72.8%; Score 331.8; DB 6; Length 361;
Best Local Similarity 99.4%; Pred. No. 2.2e-95;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-3

  Query Match             55.8%;  Score 254.4;  DB 23;  Length 456;
  Best Local Similarity   72.4%;  Pred. No. 1.8e-70;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  |  ||    | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  ||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match             55.8%;  Score 254.4;  DB 17;  Length 477;
  Best Local Similarity   72.4%;  Pred. No. 1.8e-70;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||  |||||||||||    | |||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||      |  |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  |||||||||| ||||||||   || |||| | |||||||||  ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  ||||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | |||  |||| || |||    || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || |||||||||||| |||||    ||| |||| | || |  ||     || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  |||| |||  | |||||  || || |  || |  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  |||||||  | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


Search completed: March 19, 2004, 03:01:34
Job time : 2652.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	254.4	55.8	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	210.4	46.1	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33.4	7.3	3990	5	US-09-830-230A-629	Sequence 629, App
	4	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
	5	31.4	6.9	1588	6	US-10-767-701-14522	Sequence 14522, A
c	6	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
c	7	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
	8	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
	9	29.4	6.4	1355	6	US-10-045-674A-593	Sequence 593, App
c	10	29.4	6.4	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	11	29.4	6.4	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	12	29.4	6.4	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	13	29.4	6.4	1400	7	US-60-545-213-6407	Sequence 6407, Ap
c	14	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	15	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
c	16	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	17	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	18	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	19	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
	20	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	21	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	22	28.2	6.2	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	23	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
c	24	28.2	6.2	3416	6	US-10-100-683-3397	Sequence 3397, Ap
c	25	28.2	6.2	8206	6	US-10-021-698A-3529	Sequence 3529, Ap
	26	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
c	27	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
c	28	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
	29	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
	30	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
c	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	32	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
	33	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
	34	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	35	27.6	6.1	3035	6	US-10-786-892-279	Sequence 279, App
	36	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	37	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	38	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	39	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	40	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
	41	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
	42	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
	43	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
	44	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli
c	45	27.2	6.0	2880	6	US-10-100-683-5057	Sequence 5057, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 55.8%; Score 254.4; DB 5; Length 675;
Best Local Similarity 72.4%; Pred. No. 2.6e-77;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||    ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 46.1%; Score 210.4; DB 5; Length 396;
Best Local Similarity 70.7%; Pred. No. 2.3e-62;
Matches 280; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||     || ||||| |||| |  ||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||    ||    ||||| || || |||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
```

```
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
; LENGTH: 3990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1139)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1148)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1210)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1244)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1247)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1250)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1251)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

Query Match 7.3%; Score 33.4; DB 5; Length 3990;
Best Local Similarity 47.4%; Pred. No. 0.25;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy       157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db      3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy       217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db      3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy       277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db      3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 3612

Qy       337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
Db      3613 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 3643
```

RESULT 4
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
; APPLICANT: Neubauer, Antonie
; APPLICANT: Ziegler, Christina
; TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
; FILE REFERENCE: 1/1372
; CURRENT APPLICATION NUMBER: US/10/624,149A
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 60/403,282
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: DE 10233064
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: DE 10317008
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
; LENGTH: 150223
; TYPE: DNA
; ORGANISM: Equine herpesvirus 1
; PUBLICATION INFORMATION:
; AUTHORS: Telford,E.A.
; AUTHORS: Watson, M.S.
; AUTHORS: McBride, K.
; AUTHORS: Davison, A.J.
; TITLE: The DNA sequence of equine herpesvirus-1
; JOURNAL: Virology
; VOLUME: 189
; ISSUE: 1
; PAGES: 304-316
; DATE: JUL-1992
; DATABASE ACCESSION NUMBER: NC 001491, NCBI
; DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 2.9;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy       173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
Db     95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy       233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
Db     95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy       293 GAAATAATGCCACCATCGACCAGTGG 318
Db     95662 GAGAGCTTATTCACAACGAGCTGTTG 95687
```

RESULT 5
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
; LENGTH: 1588
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

Query Match 6.9%; Score 31.4; DB 6; Length 1588;
Best Local Similarity 44.5%; Pred. No. 0.79;
Matches 125; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

```
Qy        40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
Db       455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy       100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
Db       515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy       160 AACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
Db       575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy       220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
Db       635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy       280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAA 320
Db       695 GCTGGTAGCTTTGGCGGTGGTTCCACTGGCGCCGAGAAGGA 735
```

RESULT 6
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.3;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629
Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583
```

RESULT 7
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
; LENGTH: 24900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(24900)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

Query Match 6.5%; Score 29.6; DB 6; Length 24900;
Best Local Similarity 52.4%; Pred. No. 12;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989
Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
Db       9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929
Qy        327 AAAC 330
Db       9928 AATC 9925
```

RESULT 8
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

Query Match 6.4%; Score 29.4; DB 6; Length 300;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182
Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233
```

RESULT 9
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
US-10-045-674A-593

Query Match 6.4%; Score 29.4; DB 6; Length 1355;
Best Local Similarity 25.2%; Pred. No. 3.6;
Matches 39; Conservative 50; Mismatches 66; Indels 0; Gaps 0;

```
Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827
Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887
Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYG 922
```

RESULT 10
US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18

```
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2134
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2134

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 3.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   |  | || || |   |||| || ||   || |   ||||
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
                |  ||   ||    |     | || ||||    | | |     |        ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
                |||  |     | | ||  ||    ||  |   | ||  | |    |  | ||| |
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 11
US-60-545-213-2135/c
; Sequence 2135, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2135
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2135

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 3.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   |  | || || |   |||| || ||   || |   ||||
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
                |  ||   ||    |     | || ||||    | | |     |        ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
                |||  |     | | ||  ||    ||  |   | ||  | |    |  | ||| |
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 12
US-60-545-213-6406/c
; Sequence 6406, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6406
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6406

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 3.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   |  | || || |   |||| || ||   || |   ||||
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
                |  ||   ||    |     | || ||||    | | |     |        ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
                |||  |     | | ||  ||    ||  |   | ||  | |    |  | ||| |
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 13
US-60-545-213-6407/c
; Sequence 6407, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6407
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6407

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 3.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   |  | || || |   |||| || ||   || |   ||||
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
                |  ||   ||    |     | || ||||    | | |     |        ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
                |||  |     | | ||  ||    ||  |   | ||  | |    |  | ||| |
```

```
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199

RESULT 14
US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-117

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 249;
  Best Local Similarity   52.5%;  Pred. No. 1.9;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||      || || ||| |   | |  | |   || | | |||
Db        180 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 121

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||     |  |  ||  || || || ||| |  | |||    |||| | | |    ||| |
Db        120 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 61

Qy        415 CA 416
              |
Db         60 GA 59


RESULT 15
US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 757
;   TYPE: DNA
;   ORGANISM: E. Coli
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(757)
;   OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 757;
  Best Local Similarity   52.5%;  Pred. No. 3.2;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||      || || ||| |   | |  | |   || | | |||
Db        264 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 323

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||     |  |  ||  || || || ||| |  | |||    |||| | | |    ||| |
Db        324 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 383

Qy        415 CA 416
              |
Db        384 GA 385


Search completed: March 19, 2004, 03:09:55
Job time : 57.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	255	55.9	680	12	BJ618688	BJ618688 BJ618688
c	2	37.4	8.2	484	9	AV430994	AV430994 AV430994
	3	35.2	7.7	813	28	BZ204853	BZ204853 CH230-390
	4	34.6	7.6	702	12	BG441624	BG441624 GA__Ea001
	5	34.6	7.6	937	12	BG444255	BG444255 GA__Ea002
	6	34.2	7.5	296	9	AV058630	AV058630 AV058630
	7	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
	8	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	9	34	7.5	234	10	BB581987	BB581987 BB581987
c	10	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	11	34	7.5	811	28	BZ469000	BZ469000 BOOAO58TF
c	12	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	13	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	14	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
	15	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	16	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	17	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	18	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
c	19	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	20	33.4	7.3	505	6	AL809250	Al809250 Triticum
	21	33.4	7.3	679	29	CC961526	CC961526 BOIGS10TF
c	22	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
c	23	33.4	7.3	699	12	BI959328	BI959328 HVSMEn001
c	24	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF
c	25	33.4	7.3	709	28	BH690914	BH690914 BOHXR64TF
	26	33.4	7.3	715	13	BU444683	BU444683 603765715
c	27	33.4	7.3	785	28	BH578793	BH578793 BOGLY20TR
c	28	33.4	7.3	815	28	BH461716	BH461716 BOHIV03TR
c	29	33.4	7.3	827	28	BH438448	BH438448 BOGVQ28TR
c	30	33.4	7.3	830	28	BH715216	BH715216 BOMIM42TR
c	31	33.4	7.3	835	28	BH575194	BH575194 BOHJO27TF
c	32	33.4	7.3	853	28	BH578402	BH578402 BOHNV61TF
c	33	33.4	7.3	1026	28	BZ458784	BZ458784 BONHB14TF
	34	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
	35	33.2	7.3	625	10	BF501499	BF501499 AT16857.5
	36	33.2	7.3	629	9	AU295325	AU295325 AU295325
	37	33.2	7.3	788	9	AU139321	AU139321 AU139321
c	38	33.2	7.3	1646	10	BF579120	BF579120 602096054
	39	33	7.2	456	12	BJ331339	BJ331339 BJ331339
	40	33	7.2	524	12	BJ367372	BJ367372 BJ367372
	41	33	7.2	547	12	BJ333958	BJ333958 BJ333958
	42	33	7.2	553	12	BJ336903	BJ336903 BJ336903
	43	33	7.2	569	12	BJ334047	BJ334047 BJ334047
	44	33	7.2	574	12	BJ330328	BJ330328 BJ330328
	45	33	7.2	607	12	BJ365766	BJ365766 BJ365766

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 55.9%; Score 255; DB 12; Length 680;
Best Local Similarity 72.4%; Pred. No. 3.8e-62;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||||| || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
AV430994/c
LOCUS AV430994 484 bp mRNA linear EST 23-AUG-2000
DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone PL027e10_r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
REFERENCE 1 (bases 1 to 484)
AUTHORS Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and Tabata,S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
PUBMED 10907854
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PL027e10_r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 9.5;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |   | |  | | ||  ||| |||||   ||| ||||| ||||||  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
              | ||  ||   ||  |||| ||| ||||||     |  ||    |||  |  || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323
```

RESULT 3
BZ204853
LOCUS BZ204853 813 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-390D21, genomic survey sequence.
ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 813)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-390D21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 390 row: D column: 21
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .813
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"

```
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             7.7%;  Score 35.2;  DB 28;  Length 813;
  Best Local Similarity   49.0%;  Pred. No. 51;
  Matches   94;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              ||| |   | | | | || |   |   ||    ||  | ||  |  |   ||  |  ||
Db        548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
                || |   || ||       || |  | |  ||||  |     |||| ||  |   |
Db        608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              ||   ||  |||||||| ||| |    | |||||  | |   | |||| ||| ||     |
Db        668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAAGGACAC 727

Qy        358 GGTAATAACGCC 369
              ||  | | | |
Db        728 TGTGGTTAGGAC 739


RESULT 4
BG441624
LOCUS       BG441624                 702 bp    mRNA    linear   EST 15-MAR-2001
DEFINITION  GA__Ea0014A14f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0014A14f, mRNA sequence.
ACCESSION   BG441624
VERSION     BG441624.1  GI:13351276
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 696.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0014A14f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 702;
  Best Local Similarity   46.5%;  Pred. No. 71;
  Matches  112;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | ||| | | ||  |  |   ||   ||  |   |  |   ||   | |  | || ||
Db        185 AACAATGACGGGAATGCGAATGGGAATGTTGGCAGCACAGACAATCAGAGCAACACCAAT 244

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |      |||  || |   |  || | ||  |    | || | ||  ||    |||
Db        245 GACAATGTAAATGCAAATGAAGATGTTACTCAAAAGGAAAACAACGATTCAAGTTCAAAT 304

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |  |     |  ||| | ||| | || |||   || |  | |  |   || | |  |||
Db        305 GAAAATGCTGCACAGAGCAACACCAACAACAATGAAAACAATAACAATGAAAATGCGGGT 364

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |  |||     |   |  | ||||       ||||   ||  | ||  ||| |||| |
Db        365 CAGGACGGGAAGAATGCCACTCAGAGTAACACTGACAACAATGAAAATGTGGGTCAGAAT 424

Qy        421 G 421
              |
Db        425 G 425


RESULT 5
BG444255
LOCUS       BG444255                 937 bp    mRNA    linear   EST 15-MAR-2001
DEFINITION  GA__Ea0023M14f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0023M14f, mRNA sequence.
ACCESSION   BG444255
VERSION     BG444255.1  GI:13353907
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 937)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 804.
FEATURES             Location/Qualifiers
     source          1. .937
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0023M14f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 937;
  Best Local Similarity   46.5%;  Pred. No. 81;
  Matches  112;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | ||| | | ||  |  |   ||   ||  |   |  |   ||   | |  | || ||
Db        185 AACAATGACGGGAATGCGAATGGGAATGTTGGCAGCACAGACAATCAGAGCAACACCAAT 244

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
```

```
Db        245 GACAATGTAAATGCAAATGAAGATGTTACTCAAAAGGAAAACAACGATTCAAGTTCAAAT 304

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        305 GAAAATGCTGCACAGAGCAACACCAACAACAATGAAAACAATAACAATGAAAATGCGGGT 364

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        365 CAGGACGGGAAGAATGCCACTCAGAGTAACACTGACAACAATGAAAATGTGGGTCAGAAT 424

Qy        421 G 421

Db        425 G 425
```

RESULT 6
AV058630
LOCUS AV058630 296 bp mRNA linear EST 23-JUN-1999
DEFINITION AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA clone 1810054H02, mRNA sequence.
ACCESSION AV058630
VERSION AV058630.1 GI:5158377
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 296)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"
/clone_lib="Mus musculus pancreas C57BL/6J adult"
ORIGIN

Query Match 7.5%; Score 34.2; DB 9; Length 296;
Best Local Similarity 53.3%; Pred. No. 63;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCG 310

Db         13 GTGGAGACCACAGTTTTTTCTAGGTCACCAAGGCTGGGTAAAGAAAAGAAGCATCAACTG 72
```

```
Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370

Db         73 TCCAGTGAAGCACAAAAAATGAAGAAGAAAATGATTACCCAAAAGTTGGAATTAAGGTAG 132

Qy        371 CGCTGGTTAATCAGA 385

Db        133 AGATGCCAAAATAGA 147
```

RESULT 7
BJ334624
LOCUS BJ334624 523 bp mRNA linear EST 05-MAR-2002
DEFINITION BJ334624 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda47o06 5', mRNA sequence.
ACCESSION BJ334624
VERSION BJ334624.1 GI:19164754
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 523)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .523
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47o06"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 523;
Best Local Similarity 57.1%; Pred. No. 82;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297

Db        398 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGANGACAATAAT 457

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACT 342

Db        458 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAAT 502
```

RESULT 8
BG671547
LOCUS BG671547 1036 bp mRNA linear EST 30-APR-2001
DEFINITION DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5', mRNA sequence.
ACCESSION BG671547
VERSION BG671547.1 GI:13893646
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1036)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C.,

```
            Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D.,
            Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
   PUBMED   12060780
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1036
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNBUB05"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /clone_lib="Rat DRG Library"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 1036;
  Best Local Similarity   49.2%;  Pred. No. 1.1e+02;
  Matches   90;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy          11 TAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTC 70
               | ||||   ||  | |  |||    |||  || | | || |||| |     |||||  |
Db         168 TGAAAGATACATGAGTGACAGTGTCCGTGATTGCAGACATTGCTGTATGATGCGTCCTGT 227

Qy          71 CACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAA 130
               |   | | | |  ||  ||||| ||  ||   | |   || | |   |   ||     |
Db         228 CTGTAGGAGTCCTCGCTGGTAACGATGCCGAGAGTGCACATCGATATCGGTGGGCGGTTA 287

Qy         131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
               | |       |  |  ||   | |    | ||| |||   ||| ||||  ||| || ||
Db         288 CATCTGAAGTGGAAGAGGTGGCTCTTTCGAACTCTGACTTTGATCTGTCAAGGGACGCTT 347

Qy         191 GTA 193
               |||
Db         348 GTA 350


RESULT 9
BB581987/c
LOCUS       BB581987                 234 bp    mRNA    linear   EST 30-NOV-2000
DEFINITION  BB581987 RIKEN full-length enriched, adult male colon Mus musculus
            cDNA clone 9030005O05 5', mRNA sequence.
ACCESSION   BB581987
VERSION     BB581987.1  GI:11478531
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
            Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
            Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
            Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
            Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
            Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9030005O05"
                     /sex="male"
                     /tissue_type="colon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male colon"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 10;  Length 234;
  Best Local Similarity   51.3%;  Pred. No. 65;
  Matches   79;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          35 TCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATC 94
               | ||  ||  | |     | |   || ||  || |    ||   || | || || ||   |
Db         200 TAGTTTTTGGTAGGTCGGTTTATGCGGGTTTCTTCTTCCACCAGGCCCGCCGCCGTTTCC 141
```

```
Qy         95 ATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTT 154
Db        140 CGAACGCCGGGTGCTTTATAGCCTTGTTCAAAATATGGTAGCGCATTTACCACCACCTTT 81

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
Db         80 CCGGTTACCAAGCCCGTGCCCGCCATTGCGCGGC 47
```

RESULT 10
CA711477/c
LOCUS CA711477 500 bp mRNA linear EST 26-NOV-2002
DEFINITION wdk2c.pk014.o3 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.o3 5' end, mRNA sequence.
ACCESSION CA711477
VERSION CA711477.1 GI:25433270
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 500)
AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE DuPont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES Location/Qualifiers
source 1. .500
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wdk2c.pk014.o3"
/tissue_type="kernel"
/clone_lib="wdk2c"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: XhoI; Wheat (Triticum aestivum L.) developing kernel, 7 days after anthesis."
ORIGIN

Query Match 7.5%; Score 34; DB 14; Length 500;
Best Local Similarity 54.5%; Pred. No. 91;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
Db        274 AGNTGNTGCAGTCGCAGCCNCCGGTCTCCCCGGCCGTNTCGAACCCGCCGGAGCTCCCCN 215

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGAC 126
Db        214 TGTGGGAGGCGGTGGGCATGACCATGGCGGCGGCGCCGGCGGTGGCCNCCAC 163
```

RESULT 11
BZ469000/c
LOCUS BZ469000 811 bp DNA linear GSS 13-DEC-2002
DEFINITION BOOAO58TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOAO58, genomic survey sequence.
ACCESSION BZ469000
VERSION BZ469000.1 GI:26764546
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 811)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOOAO58TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .811
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOOAO58"
/clone_lib="BO_1.6_2_KB_tot"
/note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match 7.5%; Score 34; DB 28; Length 811;
Best Local Similarity 48.9%; Pred. No. 1.1e+02;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```
Qy        176 TGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCG 235
Db        610 TGTTAGGCAAGCATTGTAAAACTGTGTTTCCCATTTCATCTACTGTTTATGACAAATGCT 551

Qy        236 CCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
Db        550 TTGATCTAATTCACTCTGATGTTTGGACTGCTCCTTGCTTGTCTAGGGATAGTTTTAAAT 491

Qy        296 ATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 355
Db        490 ACTTTGTCACATTCATAAATGAAAAATCGAAATACACCTGGATAACACTTATTCAAACCA 431

Qy        356 GCGGTA 361
Db        430 AAGATA 425
```

RESULT 12
AA746477/c
LOCUS AA746477 365 bp mRNA linear EST 27-JAN-1998
DEFINITION nw62e02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251194 3' similar to TR:O00555 O00555 NEURONAL CALCIUM CHANNEL ALPHA 1A SUBUNIT ISOFORM A-1. ;contains element TAR1 repetitive element ;, mRNA sequence.
ACCESSION AA746477
VERSION AA746477.1 GI:2786463
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 365)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

```
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html
           Insert Length: 661    Std Error: 0.00
           Seq primer: -40m13 fwd. ET from Amersham
           High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1251194"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 365;
  Best Local Similarity   56.9%;  Pred. No. 91;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  ||||||   ||||||   | ||   | ||| ||| | ||| |||   |  ||||
Db        119 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGTGGCCAGGCCGGCCGGGCGCCACC 60

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |  || |  ||| ||||  ||   | ||   | | |   | | |||| |
Db         59 AGCGGCCCTCGGAGGTACCCAGGCCCGCACGGCCGACGCTACTGGCCGG 11


RESULT 13
CD374421/c
LOCUS       CD374421                558 bp    mRNA    linear   EST 19-AUG-2003
DEFINITION  TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)
            Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION   CD374421
VERSION     CD374421.1  GI:31229939
KEYWORDS    EST.
SOURCE      Wuchereria bancrofti
  ORGANISM  Wuchereria bancrofti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Wuchereria.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
  TITLE     Toward the characterization of the Wuchereria bancrofti
            microfilarial transcriptome with comparisons to those of Brugia
            malayi and Onchocerca volvulus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Thomas B. Nutman
            Laboratory of Parasitic Diseases
            NIH
            Building 4 Room 126, Bethesda, MD 20892-0425, USA
            Email: tnutman@niaid.nih.gov
            Seq primer: pBluescript T3.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Wuchereria bancrofti"
                     /mol_type="mRNA"
                     /db_xref="taxon:6293"
                     /clone="TNWbmfC2H3"
                     /dev_stage="microfilaria"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="Wuchereria bancrofti microfilaria cDNA
                     (SAW95SjL-WbMf)"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 85,000 microfilariae
                     isolated from the blood of an infected individual from
                     Guyana, South America and converted to double-stranded
                     cDNA using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNA pol I. The library has 2.9 x 10E5
                     independent recombinants and the average insert size is
                     ~1kb. The library was constructed by Sandra J. Laney. The
                     library is available from Dr.S.A.Williams, email:
                     genome@smith.edu."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 558;
  Best Local Similarity   60.7%;  Pred. No. 1.1e+02;
  Matches   65;  Conservative    0;  Mismatches   40;  Indels    2;  Gaps    1;

Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              ||| |||| |||||||| |     || ||||| |||| |    | ||  | |
Db        161 TGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCCCAAATTCTANGCT 102

Qy        136 AGCATTTATC--AGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||| |  ||  ||||  ||  | | |||| |||||||||  ||
Db        101 GACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55


RESULT 14
BU765156/c
LOCUS       BU765156                559 bp    mRNA    linear   EST 10-OCT-2002
DEFINITION  sas23f11.y1 Gm-c1080 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1080-4869 5' similar to SW:D100_ARATH Q00874
            DNA-DAMAGE-REPAIR/TOLERATION PROTEIN DRT100 PRECURSOR ;, mRNA
            sequence.
ACCESSION   BU765156
VERSION     BU765156.1  GI:23733884
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Glycine max"
```

/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1080-4869"
/tissue_type="Roots of 8 day old 'Bragg' supernodulating mutant NTS382 seedlings"
/dev_stage="8 days old"
/lab_host="DH10B"
/clone_lib="Gm-c1080"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The mRNA was isolated from roots of 8 day old 'Bragg' supernodulating mutant NTS382 seedlings that were infected with Bradyhizobium japonicum, strain USDA 110, 72 hours prior to harvest. Dr. Gary Stacey generously donated the tissue. The roots were flash-frozen in liquid nitrogen. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An 'anchor' nucleotide (V=A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first- strand synthesis primer was then restricted by digestion with XhoI from Promega (40U/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using Sephacryl S-500 High Resolution (Pharmacia Biotech) in a 2-mm diameter column and a bed volume of approximately 1ml. The column eluent was precipitated, redissolved, and ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). This library was constucted in the laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at Northern Arizona University."

ORIGIN

Query Match 7.4%; Score 33.8; DB 13; Length 559;
Best Local Similarity 55.6%; Pred. No. 1.1e+02;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              |||||  || |  |    |||| | ||  |||||| ||| | |  || |||     ||||
Db        233 TGGGGTCGCAGGCGACGCCATACCAGCTCCGGCAACAGTACAGTGCCCGACCATGTGTTG 174

Qy        136 AGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGT 192
              |  ||     |||| |||||    |  || ||  |        | || | |||||||
Db        173 AAGATGCCCAAGTAGGGTTCGATCAGGGCAGCTTTGAAGGCCAAGAGTGCTGCCCGT 117
```

RESULT 15
CG178280
LOCUS CG178280 941 bp DNA linear GSS 21-AUG-2003
DEFINITION PUJDF79TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0641N14, genomic survey sequence.
ACCESSION CG178280
VERSION CG178280.1 GI:34069341
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 941)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUJDF79TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .941
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa0641N14"
/clone_lib="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high CoT selected genomic DNA library"
ORIGIN

Query Match 7.4%; Score 33.8; DB 29; Length 941;
Best Local Similarity 51.0%; Pred. No. 1.4e+02;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | ||||  | | || || | || || | || ||  | ||  ||    | | |||| |
Db        236 TTGCAGTCTACACATCAGTGGTCGTATATGCCACAGTTCCTGCACCGGCCTTTCCGCCGC 295

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGA 136
              || |   |  | ||  ||| ||| | || |   |     |  ||| ||||| ||   ||
Db        296 GGAGTACCCTCCGTCGTCACCACGACCGCCGGTCTTCGACTTGCCGGGACTGTACCCCGA 355

Qy        137 GCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGC 173
              || ||   || | ||  |     |  | || ||
Db        356 CTTCTTCTCGCCACCGCTCAACGGCGACGACTCTGGC 392
```

Search completed: March 18, 2004, 20:11:06
Job time : 2038.55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	3	AAC64619	Aac64619 Escherich
2	456	100.0	456	9	ACF36153	Acf36153 E. coli C
3	437.2	95.9	646	2	AAQ62647	Aaq62647 Fibronect
4	355.4	77.9	369	2	AAQ62646	Aaq62646 FNB curli
5	254.4	55.8	456	2	AAQ87467	Aaq87467 AgfA sequ
6	254.4	55.8	456	2	AAT74142	Aat74142 Salmonell
7	254.4	55.8	456	3	AAC64617	Aac64617 Salmonell
8	238.4	52.3	456	3	AAC64628	Aac64628 AgfA::PT3
9	236.8	51.9	456	3	AAC64626	Aac64626 AgfA::PT3
10	230.4	50.5	456	3	AAC64625	Aac64625 AgfA::PT3
11	230.4	50.5	456	3	AAC64627	Aac64627 AgfA::PT3
12	228.8	50.2	456	3	AAC64630	Aac64630 AgfA::PT3
13	227.2	49.8	456	3	AAC64631	Aac64631 AgfA::PT3
14	225.6	49.5	456	3	AAC64623	Aac64623 AgfA::PT3
15	222.4	48.8	456	3	AAC64624	Aac64624 AgfA::PT3
16	220.8	48.4	456	3	AAC64629	Aac64629 AgfA::PT3
17	216	47.4	456	3	AAC64622	Aac64622 AgfA::PT3
18	173.4	38.0	361	2	AAQ73066	Aaq73066 AgfA sequ
19	173.4	38.0	361	2	AAT74141	Aat74141 Salmonell
20	100	21.9	100	7	ACD68807	Acd68807 E. coli K
21	100	21.9	100	7	ACD68808	Acd68808 E. coli K
22	100	21.9	100	7	ACD68809	Acd68809 E. coli K
23	78	17.1	78	9	ACF36151	Acf36151 E. coli C
24	72	15.8	72	9	ACF36148	Acf36148 E. coli C
25	66	14.5	66	9	ACF36150	Acf36150 E. coli C
26	43.2	9.5	1083	5	AAS76745	Aas76745 DNA encod
27	40.2	8.8	4018	2	AAQ63879	Aaq63879 Polyhydro
28	39	8.6	1390	9	ADC86984	Adc86984 Human GPC
c 29	39	8.6	2730	5	AAS87782	Aas87782 DNA encod
c 30	37.4	8.2	1972	9	ADC87176	Adc87176 Human GPC
31	37	8.1	2580	5	AAH66822	Aah66822 C glutami
c 32	37	8.1	349980	5	AAH68529	Aah68529 C glutami
33	36	7.9	36	9	ACF36152	Acf36152 E. coli V
34	36	7.9	36	9	ACF36149	Acf36149 E. coli C
c 35	35.8	7.9	3984	4	ABL27701	Abl27701 Drosophil
36	35.8	7.9	12775	4	ABL27700	Abl27700 Drosophil
37	34.8	7.6	1200	5	AAS92828	Aas92828 DNA encod
38	34.6	7.6	1633	9	ADC87022	Adc87022 Human GPC
c 39	34.6	7.6	3893	6	ABK52154	Abk52154 Haemophil
40	34.6	7.6	155074	6	ABN85735	Abn85735 Human gen
c 41	34.2	7.5	1543	5	AAS92206	Aas92206 DNA encod
42	34.2	7.5	2000	7	ADA71938	Ada71938 Rice gene
43	34.2	7.5	3532	9	ADC87116	Adc87116 Human GPC
44	34.2	7.5	38918	9	ADC87242	Adc87242 Human GPC
c 45	34	7.5	570	4	ABA36458	Aba36458 Probe #14

ALIGNMENTS

RESULT 1
AAC64619
ID AAC64619 standard; DNA; 456 BP.
XX
AC AAC64619;
XX
DT 26-FEB-2001 (first entry)
XX
DE Escherichia coli CsgA DNA sequence SEQ ID NO:3.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Escherichia coli.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36343.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

```
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 122 G; 111 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 5.1e-134;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 2
ACF36153
ID   ACF36153 standard; DNA; 456 BP.
XX
AC   ACF36153;
XX
DT   04-DEC-2003  (first entry)
XX
DE   E. coli CsgA subunit 15 kDa protein encoding DNA.
XX
KW   Plasma protein; immune response; antibacterial; vaccine; gene therapy;
KW   gene; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag=  a
FT                   /note= "15 kDa protein"
XX
PN   WO2003064446-A2.
XX
PD   07-AUG-2003.
XX
PF   30-JAN-2003; 2003WO-EP000943.
XX
PR   31-JAN-2002; 2002GB-00002275.
XX
PA   (HANS-) HANSA MEDICAL RES AB.
XX
PI   Bjoerck L,  Olsen A,  Wikstroem M,  Herwald H;
XX
DR   WPI; 2003-646136/61.
DR   P-PSDB; ABR82651.
XX
PT   New isolated peptide capable of binding a mammalian plasma protein,
PT   useful in the manufacture of a medicament for the prevention and/or
PT   treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT   or Shigella infections.
XX
PS   Disclosure; Page 41-42; 42pp; English.
XX
CC   The invention relates to an isolated peptide capable of binding a
CC   mammalian plasma protein or of generating an immune response in a mammal
CC   selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC   antibody is useful for treating a bacterial infection in a human or
CC   animal or in the manufacture of a medicament for the prophylactic
CC   treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC   or Shigella infection. The peptide that is immobilized on a solid support
CC   is also useful as a reagent for determining the ability of a plasma
CC   protein to bind to bacteria. The present sequence represents an E. coli
CC   15 kDa protein encoding DNA
XX
SQ   Sequence 456 BP; 113 A; 110 C; 122 G; 111 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 456;  DB 9;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 5.1e-134;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

```
RESULT 3
AAQ62647
ID   AAQ62647 standard; DNA; 646 BP.
XX
AC   AAQ62647;
XX
DT   25-MAR-2003  (revised)
DT   18-OCT-1994  (first entry)
XX
DE   Fibronectin binding protein fnbA gene.
XX
KW   Fibronectin binding protein; FBP; curlin; curli pili; pilus;
KW   immunization; vaccine; fnbA gene; ss.
XX
OS   Escherichia coli.
XX
PN   WO9410330-A1.
XX
PD   11-MAY-1994.
XX
PF   03-NOV-1993;   93WO-US010547.
XX
PR   03-NOV-1992;   92US-00970846.
XX
PA   (REAT/) REA T S.
XX
PI   Normark S,  Olsen A;
XX
DR   WPI; 1994-167484/20.
DR   P-PSDB; AAR52664.
XX
PT   New fibronectin binding protein curlin from E. coli - used for the
PT   treatment of infections caused by E. coli and for immunisation against E.
PT   coli infection.
XX
PS   Disclosure; Fig 2; 33pp; English.
XX
CC   The fnbA gene of E. coli cattle fecal isolate AO12 includes DNA encoding
CC   a new fibronectin binding protein (FBP), curlin, which is produced as
CC   curli pili when expressed in E. coli (pFnb20) DSM 4585. The fnbA gene
CC   sequence from pFnb20 is given in AAQ62647, and includes the sequence
CC   encoding the 17 kDa curlin subunit (amino acids 21-142 of AAR52664).
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 646 BP; 168 A; 146 C; 161 G; 171 T; 0 U; 0 Other;

  Query Match           95.9%;  Score 437.2;  DB 2;  Length 646;
  Best Local Similarity  98.9%;  Pred. No. 5.4e-128;
  Matches  451;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGAGCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTC--TCGTCAACGTGACTCAGGTT 500

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||
Db        501 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 536


RESULT 4
AAQ62646
ID   AAQ62646 standard; DNA; 369 BP.
XX
AC   AAQ62646;
XX
DT   25-MAR-2003  (revised)
DT   18-OCT-1994  (first entry)
XX
DE   FNB curlin.
XX
KW   Fibronectin binding protein; FBP; curlin; curli pili; pilus;
KW   immunization; vaccine; ss.
XX
OS   Escherichia coli.
XX
PN   WO9410330-A1.
XX
PD   11-MAY-1994.
XX
PF   03-NOV-1993;   93WO-US010547.
XX
PR   03-NOV-1992;   92US-00970846.
XX
PA   (REAT/) REA T S.
XX
PI   Normark S,  Olsen A;
XX
DR   WPI; 1994-167484/20.
DR   P-PSDB; AAR52663.
XX
PT   New fibronectin binding protein curlin from E. coli - used for the
PT   treatment of infections caused by E. coli and for immunisation against E.
PT   coli infection.
XX
PS   Claim 8; Page 22; 33pp; English.
XX
CC   The fnbA gene of E. coli cattle fecal isolate AO12 includes DNA encoding
CC   a new fibronectin binding protein, curlin, which is produced as curli
CC   pili when expressed in E. coli (pFnb20) DSM 4585. (Updated on 25-MAR-2003
CC   to correct PN field.)
XX
SQ   Sequence 369 BP; 86 A; 87 C; 105 G; 91 T; 0 U; 0 Other;

  Query Match           77.9%;  Score 355.4;  DB 2;  Length 369;
  Best Local Similarity  99.2%;  Pred. No. 3.5e-102;
  Matches  368;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy         60 AGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGG 119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGG 60

Qy        120 CCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCA 120

Qy        180 AACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGA 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGA 180

Qy        240 TGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAG 299
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAG 240

Qy        300 CGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGG 359
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGG 300

Qy        360 TGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGT 419
                |||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        301 TGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTC--TCGTCAACGTGACTCAGGT 358

Qy        420 TGGCTTTGGTA 430
                ||||||||||||
Db        359 TGGCTTTGGTA 369


RESULT 5
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag=  a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             55.8%;  Score 254.4;  DB 2;  Length 456;
  Best Local Similarity   72.4%;  Pred. No. 3.9e-70;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
               || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
               ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
               |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
               || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
               || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
               || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAT74142
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
```

CC family Enterobacteria. It can also be used to provide proteins and
CC antibodies which can be used for assays. The nucleic acid sequence can be
CC used to provide probes or primers which can specifically hybridise to
CC nucleic acid molecules from greater than 99% of Salmonella strains that
CC are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC correct PF field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 55.8%; Score 254.4; DB 2; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.9e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 7
AAC64617
ID AAC64617 standard; DNA; 456 BP.
XX
AC AAC64617;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;

XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36341.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 55.8%; Score 254.4; DB 3; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.9e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 8
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match          52.3%;  Score 238.4;  DB 3;  Length 456;
  Best Local Similarity  70.2%;  Pred. No. 4.5e-65;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| | |||   |  |||||  |   | | |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |   ||    |  | ||||| |||||| | |||  ||| |  ||| |||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 9
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 51.9%; Score 236.8; DB 3; Length 456;
Best Local Similarity 70.0%; Pred. No. 1.5e-64;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| | |||   |   ||||||   ||   |  | |    ||| ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 50.5%; Score 230.4; DB 3; Length 456;
Best Local Similarity 69.1%; Pred. No. 1.6e-62;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||| | |    ||  ||      | |  |  | |    |||| ||| ||||||
```

```
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36351.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 50.5%; Score 230.4; DB 3; Length 456;
Best Local Similarity 69.1%; Pred. No. 1.6e-62;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 50.2%; Score 228.8; DB 3; Length 456;
Best Local Similarity 68.9%; Pred. No. 5e-62;
Matches 314; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || ||||                 | | ||   |   || ||  ||| | || |
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||    ||| | ||||| ||||||  | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36355.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             49.8%;  Score 227.2;  DB 3;  Length 456;
  Best Local Similarity   68.6%;  Pred. No. 1.6e-61;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||   |||||||  |      |      | |           |   |||    || |||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 225.6;  DB 3;  Length 456;
  Best Local Similarity   68.4%;  Pred. No. 5.2e-61;
  Matches  312;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||   | || ||||||||||  ||||| || || || || ||     || |  | |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                          ||  |||  |||        |  |  |   ||| |  |||  |||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 15
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 48.8%; Score 222.4; DB 3; Length 456;
Best Local Similarity 68.0%; Pred. No. 5.3e-60;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
                ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
               || ||   | |||| |   | | | |     | ||     |   ||     ||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
               ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
               |  |||||||||||| |||||   ||| |||||||||   ||     ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
               || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
               || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                  ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
               || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

Search completed: March 15, 2004, 17:51:46
Job time : 247.684 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		404.8	88.8	456	6	I44909	I44909 Sequence 58
2		404.8	88.8	2067	1	SEU43280	U43280 Salmonella
3		396.8	87.0	5103	1	STAJ2301	AJ002301 Salmonell
4		396.8	87.0	22411	1	AE008749	AE008749 Salmonell
5		395.2	86.7	254050	1	AL627269	AL627269 Salmonell
6	c	395.2	86.7	301983	1	AE016840	AE016840 Salmonell
7		389.4	85.4	1048	1	STAGFBA	AJ000514 Salmonell
8		315.2	69.1	361	6	I44908	I44908 Sequence 56
9		275.2	60.4	2889	1	CSP515700	AJ515700 Citrobact
10		232.4	51.0	2920	1	CFR515701	AJ515701 Citrobact
11		229	50.2	230	1	SEU53207	U53207 Salmonella
12		217.6	47.7	4680	1	ECCSGABDG	X90754 E.coli csgG
13		217.6	47.7	10346	1	AE000205	AE000205 Escherich
14		217.6	47.7	15047	1	D90741	D90741 Escherichia
15		216	47.4	456	6	AX814811	AX814811 Sequence
16		216	47.4	648	1	ECOCSGAA	L04979 Escherichia
17		212.6	46.6	306358	1	AE016759	AE016759 Escherich
18		209.4	45.9	1711	1	AF275733	AF275733 Escherich
19		209.4	45.9	10190	1	AE005315	AE005315 Escherich
20		209.4	45.9	327773	1	AP002554	AP002554 Escherich
21		184	40.4	2883	1	ESA515702	AJ515702 Enterobac
22		155.2	34.0	10370	1	AE015131	AE015131 Shigella
23		155.2	34.0	292504	1	AE016981	AE016981 Shigella
24	c	143.2	31.4	437	1	AF237726	AF237726 Shigella
25		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
26		54.8	12.0	19201	1	D90742	D90742 Escherichia
27		44.8	9.8	72	6	AX814798	AX814798 Sequence
28	c	39.4	8.6	2000	6	AX655393	AX655393 Sequence
29	c	38	8.3	6669	5	AF229448	AF229448 Danio rer
30	c	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
31	c	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
32	c	36	7.9	168267	2	AC146009	AC146009 Pan trogl
33		35.6	7.8	301214	1	AE016786	AE016786 Pseudomon
34	c	35	7.7	290029	1	AE017038	AE017038 Bacillus
35		34.4	7.5	547	5	AF523644	AF523644 Gavia imm
36		34.4	7.5	151828	2	AC146008	AC146008 Pan trogl
37	c	34.4	7.5	183711	2	AP001445	AP001445 Homo sapi
38	c	34.4	7.5	211981	9	AC009716	AC009716 Homo sapi
39		34.2	7.5	1458	1	AF173043	AF173043 Orientia
40		34.2	7.5	1803	3	AF325159	AF325159 Plasmodiu
41		34.2	7.5	1803	3	AF325160	AF325160 Plasmodiu
42		34.2	7.5	1803	3	AF325161	AF325161 Plasmodiu
43	c	34.2	7.5	104680	9	AC024591	AC024591 Homo sapi
44	c	34.2	7.5	189756	9	AC092327	AC092327 Homo sapi
45		34	7.5	956	8	AK058943	AK058943 Oryza sat

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match              88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity    93.0%;  Pred. No. 7e-104;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match              88.8%;  Score 404.8;  DB 1;  Length 2067;
  Best Local Similarity    93.0%;  Pred. No. 7.1e-104;
```

```
 Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYPSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match             87.0%;  Score 396.8;  DB 1;  Length 5103;
  Best Local Similarity   91.9%;  Pred. No. 1.3e-101;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |  ||  ||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES            Location/Qualifiers
     source         1. .22411
                    /organism="Salmonella typhimurium LT2"
                    /mol_type="genomic DNA"
                    /strain="LT2; SGSC 1412; ATCC 700720"
                    /db_xref="ATCC:700720"
                    /db_xref="taxon:99287"
                    /note="LT2"
     gene           434. .1308
                    /gene="phoH"
                    /note="synonym: STM1126"
     RBS            434. .439
                    /gene="phoH"
                    /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS            454. .1308
                    /gene="phoH"
                    /note="similar to E. coli PhoB-dependent, ATP-binding pho
                    regulon component; may be helicase; induced by P
                    starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                    aa), 92% identity in aa 71 - 354"
                    /codon_start=1
                    /transl_table=11
                    /product="PhoB-dependent ATP-binding pho regulon
                    component"
                    /protein_id="AAL20057.1"
                    /db_xref="GI:16419642"
                    /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                    RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                    GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                    YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                    LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                    QRTLHAYS"
     gene           complement(1414. .2303)
                    /gene="STM1127"
     CDS            complement(1414. .2295)
                    /gene="STM1127"
                    /note="similar to E. coli orf, hypothetical protein
                    (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                    identity in aa 15 - 264"
                    /codon_start=1
                    /transl_table=11
                    /product="putative transcriptional regulator"
                    /protein_id="AAL20058.1"
                    /db_xref="GI:16419643"
                    /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                    SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                    SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                    MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                    TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                    QARESKLRTMNALDMTK"
     RBS            complement(2298. .2303)
                    /gene="STM1127"
                    /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene           complement(2580. .4088)
                    /gene="STM1128"
     CDS            complement(2580. .4076)
                    /gene="STM1128"
                    /note="similar to E. coli putative cotransporter
                    (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                    identity in aa 7 - 478"
                    /codon_start=1
                    /transl_table=11
                    /product="putative sodium/glucose cotransporter"
                    /protein_id="AAL20059.1"
                    /db_xref="GI:16419644"
                    /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                    VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                    LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                    LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                    DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                    NAKLVAVIPVFFPAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                    AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                    VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                    RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS            complement(4083. .4088)
                    /gene="STM1128"
                    /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene           complement(4413. .5099)
                    /gene="STM1129"
     CDS            complement(4413. .5093)
                    /gene="STM1129"
                    /note="similar to E. coli putative enzyme (AAC76255.1);
                    Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                    226"
                    /codon_start=1
                    /transl_table=11
                    /product="putative inner membrane protein"
                    /protein_id="AAL20060.1"
                    /db_xref="GI:16419645"
                    /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                    AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                    AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                    ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                    WFSHAVKR"
     RBS            complement(5094. .5099)
                    /gene="STM1129"
                    /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene           5599. .6759
                    /gene="STM1130"
     CDS            5599. .6759
                    /gene="STM1130"
                    /note="similar to E. coli orf, hypothetical protein
                    (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                    identity in aa 40 - 403"
                    /codon_start=1
                    /transl_table=11
                    /product="putative inner membrane protein"
                    /protein_id="AAL20061.1"
                    /db_xref="GI:16419646"
                    /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                    AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                    GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                    EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                    PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                    QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                    WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene           6793. .7497
                    /gene="STM1131"
     RBS            6793. .6798
                    /gene="STM1131"
                    /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match             87.0%;  Score 396.8;  DB 1;  Length 22411;
  Best Local Similarity   91.9%;  Pred. No. 1.3e-101;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||| |     | | || | ||  |    | | | |
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
  PUBMED    11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQPGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             86.7%;  Score 395.2;  DB 1;  Length 254050;
  Best Local Similarity   91.7%;  Pred. No. 3.9e-101;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  |||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361
```

```
RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
     CDS             /locus_tag="t1735"
                     complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
     gene            REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
                     complement(844. .1797)
                     /gene="flgL"
     CDS             /locus_tag="t1736"
                     complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
     gene            NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
                     complement(1812. .3473)
                     /gene="flgK"
     CDS             /locus_tag="t1737"
                     complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
     gene            QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
                     complement(3538. .4488)
                     /gene="flgJ"
     CDS             /locus_tag="t1738"
                     complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
     gene            NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
                     complement(4488. .5591)
                     /gene="flgI"
     CDS             /locus_tag="t1739"
                     complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.7%;  Score 395.2;  DB 1;  Length 301983;
Best Local Similarity   91.7%;  Pred. No. 3.9e-101;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match           85.4%;  Score 389.4;  DB 1;  Length 1048;
  Best Local Similarity  91.0%;  Pred. No. 1.7e-99;
  Matches  414;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||||| |          | | || | ||  |    | | | |
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTA 455
              |    ||  ||||||||  ||||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           69.1%;  Score 315.2;  DB 6;  Length 361;
  Best Local Similarity  99.1%;  Pred. No. 2.1e-78;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db'        61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              |||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           60.4%;  Score 275.2;  DB 1;  Length 2889;
  Best Local Similarity  76.8%;  Pred. No. 4.9e-67;
  Matches  350;  Conservative    0;  Mismatches  103;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| |||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |     ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||   |  ||||||| |||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||   ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || ||||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||| ||||| ||  |   |  |   | | || | ||| |    ||| |  |
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
  Query Match           51.0%;  Score 232.4;  DB 1;  Length 2920;
  Best Local Similarity  71.1%;  Pred. No. 7.3e-55;
  Matches  324;  Conservative    0;  Mismatches  126;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||    ||| ||||   ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||    ||||| || ||   |||  ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||| ||| | ||||| |||   ||||  ||||  || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| ||||| |||||||| ||·|||||  | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                ||||  ||| |||||||| |       |   ||  |||  | |  ||  |  |
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || |||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

```
RESULT 11
SEU53207
LOCUS       SEU53207                 230 bp    DNA     linear   BCT 01-MAY-1996
DEFINITION  Salmonella enteritidis SEF17 fimbrin (agfA) gene, partial cds.
ACCESSION   U53207
VERSION     U53207.1  GI:1293677
KEYWORDS
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 230)
  AUTHORS   Cox,J.M., Eglezos,S. and Woolcock,J.B.
  TITLE     Virulence of Salmonella enteritidis in chickens correlates with
            colony morphology and expression of SEF17 fimbriae
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230)
  AUTHORS   Cox,J.M., Eglezos,S. and Woolcock,J.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-1996) J.M. Cox, Food Science and Technology, The
            University of New South Wales, Sydney, NSW 2052, Australia
FEATURES             Location/Qualifiers
     source          1. .230
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="SE30"
                     /db_xref="taxon:592"
                     /note="amplification product"
     gene            1. .230
                     /gene="agfA"
     CDS             <1. .>230
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="SEF17 fimbrin"
                     /protein_id="AAA98671.1"
                     /db_xref="GI:1293678"
                     /translation="GNHXGGGNSSGPDSTLSIYQYGSANAALALQSDARKSETTITQS
                     GYGNGADVGQGADNSTIELTQNGFRNNATIDQ"
ORIGIN
```

```
  Query Match           50.2%;  Score 229;  DB 1;  Length 230;
  Best Local Similarity  99.6%;  Pred. No. 6.7e-54;
  Matches  229;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         88 GGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 147
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTAATCATAANGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 60

Qy        148 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACC 207
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACC 120

Qy        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 180

Qy        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 230
```

```
RESULT 12
ECCSGABDG
LOCUS       ECCSGABDG                4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
```

```
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           47.7%;  Score 217.6;  DB 1;  Length 4680;
  Best Local Similarity  67.3%;  Pred. No. 1.2e-50;
  Matches  307;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy        1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
            ||||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db     3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy       61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
            || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db     3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy      121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
            ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db     3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy      181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
            |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db     3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy      241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
            || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db     3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy      301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
            || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db     4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy      361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||    ||| | |  |           ||  ||          | | |
Db     4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy      421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
            |    ||  |||||||| || ||| | ||||| |||
Db     4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184
```

RESULT 13
AE000205
LOCUS AE000205 10346 bp DNA linear BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 95 of 400 of the complete genome.
ACCESSION AE000205 U00096
VERSION AE000205.1 GI:1787265
KEYWORDS .
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10346)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 10346)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 10346)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 10346)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES Location/Qualifiers
source 1. .10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
repeat_region complement(<1. .48)
/note="IS3"
promoter <1. .5
/note="factor Sigma70; predicted +1 start at 1094689"
promoter 5. .32
/note="factor Sigma70; predicted +1 start at 1094716"
gene 69. .392
/gene="b1028"
CDS 69. .392
/gene="b1028"
/function="orf; Unknown"
/note="o107; This 107 aa ORF is 27 pct identical (0 gaps) to 58 residues of an approx. 1888 aa protein FAS2_CANAL SW: P43098"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74113.1"
/db_xref="GI:1787266"
/translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD AKVFK"
promoter 275. .299
/gene="b1028"
/note="factor Sigma32; predicted +1 start at 1094983"
gene 389. .1375
/gene="ycdU"
/note="synonym: b1029"
CDS 389. .1375
/gene="ycdU"
/function="orf; Unknown"
/note="o328; This 328 aa ORF is 19 pct identical (4 gaps) to 191 residues of an approx. 288 aa protein Y320_MYCGE SW: P47562"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"
/translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
protein_bind 1253. .1280
/gene="ycdU"
/note="central position to predicted promoter: -104"
/bound_moiety="DeoR predicted site"
protein_bind 1253. .1280
/gene="ycdU"
/note="No predicted promoter"
/bound_moiety="DeoR predicted site"
promoter 1337. .1365
/gene="ycdU"
/note="factor Sigma70; predicted +1 start at 1096049"
gene 1494. .1745
/gene="b1030"
CDS 1494. .1745
/gene="b1030"
/function="orf; Unknown"
/note="o83; This 83 aa ORF is 25 pct identical (7 gaps) to81 residues of an approx. 616 aa protein ALBU_RABIT SW: P49065"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC

```
                VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene       1513. .1926
                /gene="ycdV"
                /note="synonym: b1031"
     CDS        1513. .1926
                /gene="ycdV"
                /function="putative structure; Not classified"
                /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                P05759"
                /codon_start=1
                /transl_table=11
                /product="putative ribosomal protein"
                /protein_id="AAC74116.1"
                /db_xref="GI:1787269"
                /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene       complement(2111. .2198)
                /gene="serX"
                /note="synonym: b1032"
     tRNA       complement(2111. .2198)
                /gene="serX"
                /product="tRNA-Ser"
                /note="anticodon: GGA; CG Site No. 17878"
                /function="RNA; tRNA"
     promoter   2269. .2296
                /note="factor Sigma70; predicted +1 start at 1096980"
     promoter   2338. .2365
                /note="factor Sigma70; predicted +1 start at 1097049"
     promoter   2355. .2383
                /note="factor Sigma70; predicted +1 start at 1097067"
     gene       2393. .3370
                /gene="ycdW"
                /note="synonym: b1033"
     CDS        2393. .3370
                /gene="ycdW"
                /function="putative enzyme; Not classified"
                /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                to 177 residues of an approx. 312 aa protein YPRA_CORGL
                SW: P45637"
                /codon_start=1
                /transl_table=11
                /product="putative dehydrogenase"
                /protein_id="AAC74117.1"
                /db_xref="GI:1787270"
                /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter   3278. .3308
                /gene="ycdW"
                /note="factor Sigma70; predicted +1 start at 1097992"
     promoter   3383. .3410
                /note="factor Sigma70; predicted +1 start at 1098094"
     promoter   3396. .3423

  Query Match          47.7%;  Score 217.6;  DB 1;  Length 10346;
  Best Local Similarity  67.3%;  Pred. No. 1.2e-50;
  Matches  307;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || |||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||

Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                |||| ||   ||| |  |  |        ||  ||        | | |
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448


RESULT 14
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
```

```
            Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
            Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
            Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
            Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
            Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
          Headed by:
            Name: Takashi Horiuchi
            Address: National Institute of Basic Biology, Okazaki, 444, Japan
            E-mail: kishori@nibb.ac.jp
          Information operator:
            Name: Hirotada Mori
            Address: NARA Institute of Science and Technology,
                   Ikoma, 630-01, Japan
            E-mail: hmori@gtc.aist-nara.ac.jp
          URL:
            The Japan E. coli genome database
            http:bsw3.aist-nara.ac.jp.
FEATURES    Location/Qualifiers
     source 1. .15047
            /organism="Escherichia coli K12"
            /mol_type="genomic DNA"
            /strain="K12"
            /db_xref="taxon:83333"
            /map="23.7-24.0 min"
            /clone="Kohara clone #231"
            /note="Nucleotide position 1097101-1112147 from the
            initiation site of ThrA (0 min.).~This clone is from
     CDS    Kohara lambda miniset library."
            <1. .276
            /note="unnamed protein product; ORF_ID:o230#5"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA35822.1"
            /db_xref="GI:4062605"
            /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
     CDS    WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
            1152. .2090
            /note="ORF_ID:o231#1
            similar to SwissProt Accession Number P45637"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical 33.0 kd protein in proB-proA
            intergenic region."
            /protein_id="BAA35823.1"
            /db_xref="GI:4062606"
            /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
            WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
            AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
            PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
            GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
     CDS    AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
            2145. .2882
            /note="unnamed protein product; ORF_ID:o231#2"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA35824.1"
            /db_xref="GI:4062607"
            /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
            EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
            PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
            SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
     CDS    ILNVSPRRLLNFLESRGMAPIAEFADL"
            2906. .3460
            /note="ORF_ID:o231#3
            similar to PIR Accession Number F64035"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein HI1543"
            /protein_id="BAA35825.1"
            /db_xref="GI:4062608"
            /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
            DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
            ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
     CDS    RTMAPLTRDAISAMWDELEEDSEE"
            3514. .4053
            /note="unnamed protein product; ORF_ID:o231#4"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA35826.1"
            /db_xref="GI:4062609"
            /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
            SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
            IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
     gene   LAARSAKTAHREQEIKNKA"
            complement(4117. .4950)
     CDS    /gene="csgG"
            complement(4117. .4950)
            /gene="csgG"
            /note="ORF_ID:o231#5
            similar to SwissProt Accession Number P52103"
            /codon_start=1
            /transl_table=11
            /product="Curli production assembly/transport component
            CsgG precursor."
            /protein_id="BAA35827.1"
            /db_xref="GI:4062610"
            /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
            PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
            NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
            IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
            VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
     gene   S"
            complement(4977. .5393)
     CDS    /gene="csgF"
            complement(4977. .5393)
            /gene="csgF"
            /note="ORF_ID:o231#6
            similar to SwissProt Accession Number P52104"
            /codon_start=1
            /transl_table=11
            /product="Curli production assembly/transport component
            CsgF precursor."
            /protein_id="BAA35828.1"
            /db_xref="GI:4062611"
            /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
            AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
     gene   DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
            complement(5418. .5807)
     CDS    /gene="csgE"
            complement(5418. .5807)
            /gene="csgE"
            /note="ORF_ID:o231#7
            similar to SwissProt Accession Number P52105"
            /codon_start=1
            /transl_table=11
            /product="Curli production assembly/transport component
            CsgE precursor."
            /protein_id="BAA35829.1"
            /db_xref="GI:4062612"
            /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
            RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
     gene   IQTEEALNRRQINQALLSTGDLAHDEF"
            complement(5812. .6462)
     CDS    /gene="csgD"
            complement(5812. .6462)
            /gene="csgD"
            /note="ORF_ID:o231#8
            similar to SwissProt Accession Number P52106"
            /codon_start=1
            /transl_table=11
            /product="Probable csgAB operon transcriptional regulatory
            protein."
            /protein_id="BAA35830.1"
            /db_xref="GI:4062613"
            /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
```

```
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            7217. .7672
                     /gene="csgB"
     CDS             7217. .7672

  Query Match          47.7%;  Score 217.6;  DB 1;  Length 15047;
  Best Local Similarity   67.3%;  Pred. No. 1.2e-50;
  Matches  307;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||    ||| |  |   |          ||   ||          | |  |
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 15
AX814811
LOCUS       AX814811                456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match          47.4%;  Score 216;  DB 6;  Length 456;
  Best Local Similarity   67.1%;  Pred. No. 3.3e-50;
  Matches  306;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||    ||| |  |   |          ||   ||          | |  |
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


Search completed: March 18, 2004, 15:02:42
Job time : 2312.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64622	Aac64622 AgfA::PT3
	2	404.8	88.8	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	404.8	88.8	456	2	AAT74142	Aat74142 Salmonell
	4	404.8	88.8	456	3	AAC64617	Aac64617 Salmonell
	5	376	82.5	456	3	AAC64623	Aac64623 AgfA::PT3
	6	361.6	79.3	456	3	AAC64626	Aac64626 AgfA::PT3
	7	355.2	77.9	456	3	AAC64625	Aac64625 AgfA::PT3
	8	353.6	77.5	456	3	AAC64628	Aac64628 AgfA::PT3
	9	353.6	77.5	456	3	AAC64629	Aac64629 AgfA::PT3
	10	352	77.2	456	3	AAC64624	Aac64624 AgfA::PT3
	11	350.4	76.8	456	3	AAC64630	Aac64630 AgfA::PT3
	12	350.4	76.8	456	3	AAC64627	Aac64627 AgfA::PT3
	13	347.2	76.1	456	3	AAC64631	Aac64631 AgfA::PT3
	14	315.2	69.1	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	315.2	69.1	361	2	AAT74141	Aat74141 Salmonell
	16	216	47.4	456	3	AAC64619	Aac64619 Escherich
	17	216	47.4	456	9	ACF36153	Acf36153 E. coli C
	18	211.6	46.4	646	2	AAQ62647	Aaq62647 Fibronect
	19	163.2	35.8	369	2	AAQ62646	Aaq62646 FNB curli
	20	52.2	11.4	78	3	AAC64610	Aac64610 AgfA (SEF
c	21	49.4	10.8	78	3	AAC64609	Aac64609 AgfA (SEF
	22	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	23	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
	24	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
c	25	40.8	8.9	78	3	AAC64605	Aac64605 SefA (SEF
	26	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
	27	40.6	8.9	78	3	AAC64606	Aac64606 SefA (SEF
c	28	39.4	8.6	2000	7	ADA71938	Ada71938 Rice gene
	29	39	8.6	100	7	ACD68808	Acd68808 E. coli K
	30	36	7.9	557	6	ABQ43459	Abq43459 Oligonucl
c	31	36	7.9	557	6	ABQ43458	Abq43458 Oligonucl
	32	36	7.9	646	6	ABQ46953	Abq46953 Oligonucl
c	33	36	7.9	646	6	ABQ46952	Abq46952 Oligonucl
	34	35.4	7.8	100	7	ACD68807	Acd68807 E. coli K
c	35	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	36	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
	37	34.6	7.6	456	3	AAC64620	Aac64620 Escherich
	38	34	7.5	456	3	AAC64618	Aac64618 Salmonell
c	39	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	40	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	41	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	42	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	43	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	44	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c	45	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a

ALIGNMENTS

RESULT 1
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.9e-137;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy 301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db 301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy 361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
Db 361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy 421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
Db 421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 2; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.1e-120;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db 241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy 301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db 301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy 361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
Db 361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy 421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
Db 421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 2;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.1e-120;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

```
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 3;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.1e-120;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | || |  | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match             82.5%;  Score 376;  DB 3;  Length 456;
  Best Local Similarity   89.0%;  Pred. No. 2.4e-111;
  Matches  406;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
                  ||    | ||||   ||||   |      | |  | ||  |    | | |  |
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match           79.3%;  Score 361.6;  DB 3;  Length 456;
  Best Local Similarity 87.1%;  Pred. No. 1.1e-106;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| |||   |   ||||||   | |   | |   ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |   | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 1.3e-104;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |   ||   |     | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||| |          | | || | ||  |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             77.5%;  Score 353.6;  DB 3;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 4.4e-104;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              || |    || |  | ||| |          | | || | ||  |    | | |  |
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 9
AAC64629
ID AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

```
  Query Match             77.5%;  Score 353.6;  DB 3;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 4.4e-104;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 1.5e-103;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy    1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db    1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy   61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db   61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy  121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db  121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy  181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db  181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy  241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db  241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy  301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db  301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy  361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
Db  361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy  421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
Db  421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 4.8e-103;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | |||  |    |  | ||  |   |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||||  |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            76.8%;  Score 350.4;  DB 3;  Length 456;
  Best Local Similarity  85.5%;  Pred. No. 4.8e-103;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||  |  |       | ||  |    || |  |   | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||  |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP, Doran JL, Collison SK, Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match           76.1%;  Score 347.2;  DB 3;  Length 456;
  Best Local Similarity 85.1%;  Pred. No. 5.3e-102;
  Matches  388;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |   |           |   | | ||   ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||| |     | | || | ||   |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             69.1%;  Score 315.2;  DB 2;  Length 361;
  Best Local Similarity   99.1%;  Pred. No. 1.1e-91;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              |||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             69.1%;  Score 315.2;  DB 2;  Length 361;
  Best Local Similarity   99.1%;  Pred. No. 1.1e-91;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              |||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320


Search completed: March 18, 2004, 09:11:55
```

Job time : 289.405 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	404.8	88.8	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	315.2	69.1	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	33.4	7.3	7766	4	US-09-125-619-3	Sequence 3, Appli
	6	33.4	7.3	14770	4	US-09-220-132-30	Sequence 30, Appl
c	7	32.2	7.1	1122	4	US-09-489-039A-3211	Sequence 3211, Ap
c	8	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	9	31.2	6.8	2436	4	US-09-540-236-492	Sequence 492, App
c	10	31.2	6.8	65792	4	US-09-596-002-31	Sequence 31, Appl
	11	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
c	12	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	13	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	14	30	6.6	1341	4	US-09-350-756-2	Sequence 2, Appli
	15	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	16	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	17	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
c	18	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
c	19	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	20	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	21	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	22	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	23	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	24	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, App
c	25	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, App
c	26	29.4	6.4	3494	4	US-09-023-655-1006	Sequence 1006, Ap
c	27	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
	28	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
	29	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	30	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	31	29	6.4	534	4	US-09-489-039A-2195	Sequence 2195, Ap
	32	29	6.4	1200	4	US-09-134-000C-1704	Sequence 1704, Ap
	33	29	6.4	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
c	34	29	6.4	13794	4	US-08-956-171E-54	Sequence 54, Appl
c	35	28.8	6.3	654	4	US-09-252-991A-4506	Sequence 4506, Ap
	36	28.8	6.3	720	4	US-09-328-352-354	Sequence 354, App
	37	28.8	6.3	1662	1	US-08-565-386-2	Sequence 2, Appli
	38	28.8	6.3	4529	1	US-08-565-386-1	Sequence 1, Appli
c	39	28.8	6.3	8906	4	US-09-027-169-5	Sequence 5, Appli
c	40	28.6	6.3	432	4	US-09-252-991A-4473	Sequence 4473, Ap
	41	28.4	6.2	1026	3	US-07-751-891B-24	Sequence 24, Appl
	42	28.4	6.2	4188	3	US-07-751-891B-2	Sequence 2, Appli
	43	28.4	6.2	4242	3	US-07-705-490-2	Sequence 2, Appli
	44	28.4	6.2	5222	3	US-07-751-891B-23	Sequence 23, Appl
	45	28.2	6.2	357	4	US-09-489-039A-1118	Sequence 1118, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 4.5e-131;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             69.1%;  Score 315.2;  DB 1;  Length 361;
  Best Local Similarity   99.1%;  Pred. No. 6.5e-100;
  Matches  317;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| |||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              |||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320
```

```
RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
;  GENERAL INFORMATION:
;    APPLICANT:  Kim, Jihyun Francis
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;    TITLE OF INVENTION:  ERWINIA AMYLOVORA AND ITS USE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,927
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,108
;      FILING DATE:  06-AUG-1977
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1581
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
```

```
;       TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1344 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.13;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.13;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-09-125-619-3

  Query Match             7.3%;  Score 33.4;  DB 4;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 0.56;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  |||||| ||||     |    ||||      ||  ||| |  | | | |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |   |   | |||||||||  | |||     |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    |  ||| |          | |||  |||||  |      |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||        | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 6
US-09-220-132-30
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  TITLE OF INVENTION:  METHODS AND COMPOSITONS FOR THE IDENTIFICATION AND ASSESSMENT
;  TITLE OF INVENTION:  OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCE
;  FILE REFERENCE:  07334-074001
;  CURRENT APPLICATION NUMBER: US/09/220,132
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  US 60/079,303
;  PRIOR FILING DATE:  1998-03-25
;  PRIOR APPLICATION NUMBER:  US 60/068,821
;  PRIOR FILING DATE:  1997-12-24
;  NUMBER OF SEQ ID NOS:  191
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 30
;   LENGTH: 14770
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-220-132-30

  Query Match             7.3%;  Score 33.4;  DB 4;  Length 14770;
  Best Local Similarity   48.7%;  Pred. No. 0.8;
  Matches   91;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              || | ||| || |      ||  | ||||  ||    |||      |   | |||| | |
Db      10282 GGATCTGGAAAAGATAATGAATTTGGCCTTGGCCTTGATTCACCTCAGAATGAAATTGCC 10341

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
```

```
Db      10342 CAGAATGGGAACAACGACCAGTCCATCACAGAGTGTTCCATTGCCACCACAGCAGAGTTT 10401

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACC 399

Db      10402 TCTCATGACACGGATGCCACAGAGATCGACTCTCTGGATGGCTATGACCTGCAAGATGAA 10461

Qy        400 CGTGTTG 406

Db      10462 GATGATG 10468


RESULT 7
US-09-489-039A-3211/c
; Sequence 3211, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 3211
;   LENGTH: 1122
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3211

  Query Match             7.1%;  Score 32.2;  DB 4;  Length 1122;
  Best Local Similarity   61.2%;  Pred. No. 0.49;
  Matches   52;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370

Db       1007 AGCACCGGACCGATAAACAGGCCGATCATGCCGAAGGCGATCAGGCCGCCAATGACGCCG 948

Qy        371 CGCTGGTTAATTATGATCAGCTGGT 395

Db        947 GTGAGGATAAGGATCATTGGCAGGT 923


RESULT 8
5486473-3/c
;Patent No. 5486473
;     APPLICANT: FUKITA, HIROYUKI;YOSHIDA, IWAO;TAKAGI, MITSUO;
;MANABE, SADAO;FUKAI, KONOSUKE
;     TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
;     NUMBER OF SEQUENCES: 16
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/194,049
;       FILING DATE: 09-FEB-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 809,255
;       FILING DATE: 18-DEC-1991
;       APPLICATION NUMBER: 279,685
;       FILING DATE: 05-DEC-1988
;       APPLICATION NUMBER: 932,419
;       FILING DATE: 19-NOV-1986
;SEQ ID NO:3:
;       LENGTH: 1491
5486473-3

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 1491;
  Best Local Similarity   50.0%;  Pred. No. 1.3;
  Matches   78;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237

Db        204 CAAGTCACTGACCGAAGCTAGGTAACAGTAACTGCGCACATCTGCGAGGTTGGCTGCTTC 145

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297

Db        144 CATGTTCATCATTTTGACATCAATGGTTGGCTTGTCTTTTGACATTATGGTCACACAACT 85

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333

Db         84 ATCGCCTTCCAGTACCAGATCAACCCATGTAGCTCC 49


RESULT 9
US-09-540-236-492/c
; Sequence 492, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2005-001
;  CURRENT APPLICATION NUMBER: US/09/540,236
;  CURRENT FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 492
;   LENGTH: 2436
;   TYPE: DNA
;   ORGANISM: M.catarrhalis
US-09-540-236-492

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 2436;
  Best Local Similarity   53.7%;  Pred. No. 1.7;
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268

Db       1511 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 1452

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325

Db       1451 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 1392

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369

Db       1391 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 1348


RESULT 10
US-09-596-002-31/c
; Sequence 31, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:      Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:      PERL Program
; SEQ ID NO 31
;   LENGTH: 65792
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte template ID No. 6632636 31
;  PUBLICATION INFORMATION:
US-09-596-002-31

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 65792;
  Best Local Similarity   53.7%;  Pred. No. 11;
```

```
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              ||||||| |    | ||  |       | | |  | || || ||||| |||   | |||
Db       6819 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 6760

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |      | |  || ||||  |||| |||||| |||| |||| ||  ||  |
Db       6759 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 6700

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |   || ||  ||  |||| ||  ||||   |  |||
Db       6699 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 6656


RESULT 11
US-08-508-761B-1
; Sequence 1, Application US/08508761B
; Patent No. 6027920
;  GENERAL INFORMATION:
;    APPLICANT:  Joliff, Gwennael
;    APPLICANT:  Guyonvarch, Armel
;    APPLICANT:  Purification, Relano
;    APPLICANT:  Duchiron, Francis
;    APPLICANT:  Renaud, Michel
;    TITLE OF INVENTION:  System for Protein Expression and
;    TITLE OF INVENTION:  Secretion Especially in Corynebacteria
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jacobson, Price, Holman & Stern, PLLC
;      STREET:  400 Seventh St. N.W.
;      CITY:  Washington D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/508,761B
;      FILING DATE:  31-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09652
;      FILING DATE:  29-JUL-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09870
;      FILING DATE:  02-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Player, William E.
;      REGISTRATION NUMBER:  31,409
;      REFERENCE/DOCKET NUMBER:  P58525NA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 638-6666
;      TELEFAX:  (202) 393-5350
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2547 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Corynebacterium Melassecola
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  420..2390
US-08-508-761B-1

  Query Match             6.8%;  Score 30.8;  DB 3;  Length 2547;
  Best Local Similarity   58.9%;  Pred. No. 2.4;

  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
              |||| |||||  |  | | |  ||| ||| || ||||   ||||  | ||   ||  |||
Db        876 CAGAGTGGTTGGACCATTGAGACCAACATTGAGCAGTACTACGCCGATAAGAACGCCATT 935

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |||   || |  ||| ||  | | | ||
Db        936 GTTGTGCTCCCAATCGGTGGCGAGAGCTCC 965


RESULT 12
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4403765;
  Best Local Similarity   51.9%;  Pred. No. 95;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
              | || | || |||||||  || ||  |  ||   | |||||| |||| |  |   || ||
Db    1357218 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357159

Qy        105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
                ||     | | |     ||  | ||| | ||    |   ||   ||  || | |  |||
Db    1357158 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357099

Qy        165 TGCGCTTGCTCTG 177
                ||   |  |||
Db    1357098 CCCGGACGTGCTG 1357086


RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
```

```
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             6.7%;  Score 30.6;  DB 3;  Length 4411529;
  Best Local Similarity   51.9%;  Pred. No. 95;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          45 TGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGG 104
               | || | || ||||||| || ||  |  ||   | |||||| |||| |  |   || ||
Db     1357750 TTGCCGAGCGCTGGCTGACGGCGCGCGTCACACGCGCGGCGTCGGTGACGAGTCCGTCGA 1357691

Qy         105 CGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
                ||     | | |    ||  | ||| | ||   |   ||   ||  || | |  |||
Db     1357690 AAACACCGCATGCCGTAACGAGCCCACGATAAGGTGTGGCCATGTCGTATCTGGTGGCGC 1357631

Qy         165 TGCGCTTGCTCTG 177
                ||   | |||
Db     1357630 CCCGGACGTGCTG 1357618


RESULT 14
US-09-350-756-2/c
; Sequence 2, Application US/09350756
; Patent No. 6495143
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Reseach Institute for Infectious  Diseases
;  APPLICANT: John S. Lee
;  APPLICANT: Peter Pushko
;  APPLICANT: Michael D. Parker
;  APPLICANT: Jonathan F. Smith
;  APPLICANT: Mark T. Dertzbaugh
;  APPLICANT: Leonard Smith
;  TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
;  FILE REFERENCE: 003/124/SAP RIID 98-21
;  CURRENT APPLICATION NUMBER: US/09/350,756
;  CURRENT FILING DATE: 1999-07-09
;  EARLIER APPLICATION NUMBER: US 60/092,416
;  EARLIER FILING DATE: 1998-07-10
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 2
;   LENGTH: 1341
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-2

  Query Match             6.6%;  Score 30;  DB 4;  Length 1341;
  Best Local Similarity   53.4%;  Pred. No. 3.2;
  Matches   63;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         330 CTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCA 389
               || |||||||    |  | || | ||| | | |  | |  |   |  || ||  ||| |
Db         673 CTTCGATATTGGACTGAGACAGTTCGGTGTTGAAGATGGAGAAGTATTTCATCCAGATGA 614

Qy         390 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAAC 447
                |||| |||    | ||| |||    |  | | | |||    || ||  |  | ||||
Db         613 ACTGGGTACGATCGATGTCACCGTCCAGTTTGAAGATGATTTCACCGTTAGCGATAAC 556


RESULT 15
US-09-252-991A-10904
; Sequence 10904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10904
;   LENGTH: 1008
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904

  Query Match             6.5%;  Score 29.8;  DB 4;  Length 1008;
  Best Local Similarity   54.0%;  Pred. No. 3.2;
  Matches   61;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          41 TTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACG 100
               | | || | |  |  || | |  | ||| |  |    ||||  ||||||  | |     |
Db         141 TGTATGCCTGGTCGGTGACCGAGGGCGTGCAGCGCCTGGGCTTCGGCGGCGAGCCGCTGG 200

Qy         101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
               |||| |   | ||   || ||||| | |  |||||  || |     || ||| |
Db         201 GCGGGGAGGAGAGCCGCGACCCGGTCGCGGCGTTGCGCTTGATCAAGCACGAT 253


Search completed: March 18, 2004, 20:20:26
Job time : 68.4971 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	34.2	7.5	616	12	US-10-424-599-88945	Sequence 88945, A
c	2	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c	3	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c	4	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c	5	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c	6	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
	7	34	7.5	2886	9	US-09-801-368-131	Sequence 131, App
	8	33.6	7.4	2751	15	US-10-398-221-3563	Sequence 3563, Ap
	9	33.6	7.4	3532	14	US-10-017-161-1913	Sequence 1913, Ap
	10	33.6	7.4	3532	15	US-10-292-798-1569	Sequence 1569, Ap
	11	33.4	7.3	7766	12	US-10-222-566-3	Sequence 3, Appli
	12	33.4	7.3	7766	14	US-10-222-162-3	Sequence 3, Appli
	13	33.4	7.3	7766	14	US-10-143-024-3	Sequence 3, Appli
	14	33.4	7.3	14770	10	US-09-873-367C-230	Sequence 230, App
	15	33.2	7.3	1355	15	US-10-045-674-593	Sequence 593, App
c	16	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
	17	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	18	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
	19	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c	20	32.2	7.1	1110	12	US-10-282-122A-23637	Sequence 23637, A
c	21	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
	22	32	7.0	1776	12	US-10-282-122A-24385	Sequence 24385, A
c	23	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
	24	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
	25	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
c	26	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
	27	31.4	6.9	840	12	US-10-282-122A-31900	Sequence 31900, A
	28	31.4	6.9	1245	15	US-10-369-493-40803	Sequence 40803, A
c	29	31.2	6.8	2394	12	US-10-282-122A-26837	Sequence 26837, A
c	30	31	6.8	525	10	US-09-918-995-32144	Sequence 32144, A
	31	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
	32	30.8	6.8	725	9	US-09-919-580-878	Sequence 878, App
	33	30.8	6.8	2050	14	US-10-086-510-1	Sequence 1, Appli
	34	30.8	6.8	2699	15	US-10-260-238-418	Sequence 418, App
	35	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
	36	30.6	6.7	2301	9	US-09-815-242-4844	Sequence 4844, Ap
	37	30.6	6.7	2301	9	US-09-815-242-9042	Sequence 9042, Ap
c	38	30.6	6.7	75216	14	US-10-080-170-646	Sequence 646, App
c	39	30.4	6.7	392	12	US-10-282-122A-25231	Sequence 25231, A
c	40	30.4	6.7	992	12	US-10-425-114-31747	Sequence 31747, A
c	41	30.4	6.7	1024	15	US-10-323-069A-111	Sequence 111, App
c	42	30.4	6.7	1601042	15	US-10-027-632-59064	Sequence 59064, A
	43	30.2	6.6	654	12	US-10-425-114-28546	Sequence 28546, A
c	44	30.2	6.6	829	15	US-10-027-632-164999	Sequence 164999,
c	45	30.2	6.6	1203	9	US-09-738-626-3184	Sequence 3184, Ap

ALIGNMENTS

RESULT 1
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(616)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

Query Match 7.5%; Score 34.2; DB 12; Length 616;
Best Local Similarity 50.9%; Pred. No. 0.51;
Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy        101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
              | | | ||| ||  || ||| || || ||| | |    | |  | | | | | | ||
Db        556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497

Qy        161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
              | | ||   || | | |   || || | | | | | |||  ||  | | |
Db        496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437
```

```
Qy        221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              || |||||| |  |  ||| |||  ||||| ||| ||
Db        436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTA 398
```

RESULT 2
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |.     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 3
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 4
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 5
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora

US-10-010-390-6

Query Match 7.5%; Score 34; DB 14; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 6
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (EBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-441-736-5

Query Match 7.5%; Score 34; DB 15; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.93;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||| | |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 7
US-09-801-368-131
; Sequence 131, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 131
; LENGTH: 2886
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-131

Query Match 7.5%; Score 34; DB 9; Length 2886;
Best Local Similarity 50.6%; Pred. No. 1.4;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              | ||  | | | | || | | | |||||    | ||    || |   |    |||| | |
Db       1501 AACAACAACAATAACGATAACGATAACGATAACAATAATAGCAATAATAATAATAACAAT 1560

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || | |    | | | |||  |  ||  ||||||  ||  | | || |    |||  |
Db       1561 AATAATAATAACAATAATAATAATAACAATAATAATAACAATAACAACAACAACAACAAT 1620

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
              || ||    ||| || |       || |   || | |||||
Db       1621 AATAATAATAATAGTAATAATATTAAAAATAACGTTGATAAC 1662
```

RESULT 8
US-10-398-221-3563
; Sequence 3563, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3563
; LENGTH: 2751
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3563

Query Match 7.4%; Score 33.6; DB 15; Length 2751;
Best Local Similarity 51.3%; Pred. No. 1.9;
Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||||||   | |  ||| | |||| ||||||  ||      |       | || | | ||
Db        321 GATGTAGTGAAAGAAGCGAAAAATATTACTATCGACAAACCGTTTCCTCGTATGACATAT 380
```

```
Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              || |   | || || | |     |  | | ||  | |||||  | | ||   | |
Db        381 AAAGAAGCAATGGATCGTTTTGGCAGTGATAAACCAGATATTCGTTTTGGTTTAGAATTA 440

Qy        358 GGTAATAACGCCGCGCTGGTTAATTATGATCA 389
                |||    |||   | || ||  ||| | |
Db        441 CAAAATGTATCCGATGTCGTAAAAGATGTTGA 472


RESULT 9
US-10-017-161-1913
; Sequence 1913, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1913
;   LENGTH: 3532
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(3532)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(437)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2622)..(3332)
US-10-017-161-1913

  Query Match           7.4%;  Score 33.6;  DB 14;  Length 3532;
  Best Local Similarity   46.6%;  Pred. No. 2.2;
  Matches  108;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy         215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
               | |  ||| ||||||| ||      |||  |   | ||||  | | |||||| |  ||
Db        2881 ATAATGGTGATGGTAATGGTAATCATGGTGATGATGGTGGTGGTGATAGTAATCATGGTG 2940

Qy         275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
               | |  ||| |     |   | |||           |  |  |  |  |  |     |
Db        2941 ACGGTAATGATGATAGTGGTGGTGATGATGATGATGGTGATGATGATGGAGATCATGTTG 3000

Qy         335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGG 394
               || ||  ||| |       |  |||| || | |  |     | ||||  ||||||    |
Db        3001 ATGTTGATGTTGATAGTGATAACGGTGATGATGGTGGTGATGATAATGGTGATCACGGTG 3060

Qy         395 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAA 446
               ||    ||| ||||     ||| |  |   |||  ||  | | ||  | | |
Db        3061 TTGATGGTGGTGTTGATGGTGATAATGGTGATGGTAATGATGGCAATGGTGA 3112


RESULT 10
US-10-292-798-1569
; Sequence 1569, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1569
;   LENGTH: 3532
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(3532)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(437)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2622)..(3332)
US-10-292-798-1569

  Query Match           7.4%;  Score 33.6;  DB 15;  Length 3532;
  Best Local Similarity   46.6%;  Pred. No. 2.2;
  Matches  108;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy         215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
               | |  ||| ||||||| ||      |||  |   | ||||  | | |||||| |  ||
Db        2881 ATAATGGTGATGGTAATGGTAATCATGGTGATGATGGTGGTGGTGATAGTAATCATGGTG 2940

Qy         275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
               | |  ||| |     |   | |||           |  |  |  |  |  |     |
Db        2941 ACGGTAATGATGATAGTGGTGGTGATGATGATGATGGTGATGATGATGGAGATCATGTTG 3000

Qy         335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGG 394
               || ||  ||| |       |  |||| || | |  |     | ||||  ||||||    |
Db        3001 ATGTTGATGTTGATAGTGATAACGGTGATGATGGTGGTGATGATAATGGTGATCACGGTG 3060

Qy         395 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAA 446
               ||    ||| ||||     ||| |  |   |||  ||  | | ||  | | |
Db        3061 TTGATGGTGGTGTTGATGGTGATAATGGTGATGGTAATGATGGCAATGGTGA 3112


RESULT 11
US-10-222-566-3
; Sequence 3, Application US/10222566
; Publication No. US20040044192A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD3
;  CURRENT APPLICATION NUMBER: US/10/222,566
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-566-3
```

Query Match 7.3%; Score 33.4; DB 12; Length 7766;
Best Local Similarity 47.4%; Pred. No. 4;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db 4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy 217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db 4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy 277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db 4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy 337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
Db 4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642

RESULT 12
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234USD4
; CURRENT APPLICATION NUMBER: US/10/222,162
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/125,619
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-222-162-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 4;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db 4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy 217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db 4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy 277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db 4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy 337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
Db 4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642

RESULT 13
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234USD1
; CURRENT APPLICATION NUMBER: US/10/143,024
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 09/125,619
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-143-024-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 4;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db 4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy 217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db 4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy 277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db 4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy 337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
Db 4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642

RESULT 14
US-09-873-367C-230
; Sequence 230, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meena
; APPLICANT: Ebner, Reinhard
; APPLICANT: Carter, Kenneth
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; TITLE OF INVENTION: Signature Gene Sets
; FILE REFERENCE: 689290-64
; CURRENT APPLICATION NUMBER: US/09/873,367C
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,891
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,842
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/244,867
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: U.S. 60/245,084
; PRIOR FILING DATE: 2000-11-01

```
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 230
;   LENGTH: 14770
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-873-367C-230

  Query Match             7.3%;  Score 33.4;  DB 10;  Length 14770;
  Best Local Similarity   48.7%;  Pred. No. 5.7;
  Matches   91;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              || | ||| || |     ||  | |||| ||    |||      |   | |||| | | |
Db      10282 GGATCTGGAAAAGATAATGAATTTGGCCTTGGCCTTGATTCACCTCAGAATGAAATTGCC 10341

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
              ||||||||   ||   |  |  ||| ||  ||    |  |  || | |   | ||  ||
Db      10342 CAGAATGGGAACAACGACCAGTCCATCACAGAGTGTTCCATTGCCACCACAGCAGAGTTT 10401

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACC 399
               ||   | |    | | |    | | || | | |||| |   ||||| | ||  | |
Db      10402 TCTCATGACACGGATGCCACAGAGATCGACTCTCTGGATGGCTATGACCTGCAAGATGAA 10461

Qy        400 CGTGTTG 406
              || ||
Db      10462 GATGATG 10468


RESULT 15
US-10-045-674-593
; Sequence 593, Application US/10045674
; Publication No. US20030232333A1
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 60/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674-593

  Query Match             7.3%;  Score 33.2;  DB 15;  Length 1355;
  Best Local Similarity   23.9%;  Pred. No. 1.8;
  Matches   83;  Conservative   91;  Mismatches  174;  Indels    0;  Gaps    0;

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137
              :||:||:||: ::  :  :  : ::||:||:  ::::   ||:   | :::   :    |
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197
              :  :: :  :  :||::::| :: :| :|   |:| :    |:|    ||   :
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGA 257
              : |:  : |:|| ||: |:| :| : |:|  ||:|       |    | :       |   |
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        258 TAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
                :  :::: | |        : |   || |  :|   |:  :  :| :  :||:   ::
Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYATYGGYGAYGTYWS 1007

Qy        318 GAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGT 377
              :  :   ||:  : |:  : |:  :||  | | :|  ||:  :||  |:  : |||
Db       1008 YGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKAATTCYCARATGGC 1067

Qy        378 TAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 425
              : |   :| : |   :| :| : ::  : |:|   |: | |   :||
Db       1068 YCARGTYGGWGAYGGKGAYAAYWSWCCKYTWATGAAYAAYTTYMGWCA 1115


Search completed: March 19, 2004, 03:49:15
Job time : 257.734 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	23	US-09-543-407-11	Sequence 11, Appl
	2	404.8	88.8	456	6	US-08-233-642A-56	Sequence 56, Appl
	3	404.8	88.8	456	23	US-09-543-407-1	Sequence 1, Appli
	4	376	82.5	456	23	US-09-543-407-13	Sequence 13, Appl
	5	361.6	79.3	456	23	US-09-543-407-19	Sequence 19, Appl
	6	355.2	77.9	456	23	US-09-543-407-17	Sequence 17, Appl
	7	353.6	77.5	456	23	US-09-543-407-23	Sequence 23, Appl
	8	353.6	77.5	456	23	US-09-543-407-25	Sequence 25, Appl
	9	352	77.2	456	23	US-09-543-407-15	Sequence 15, Appl
	10	350.4	76.8	456	23	US-09-543-407-21	Sequence 21, Appl
	11	350.4	76.8	456	23	US-09-543-407-27	Sequence 27, Appl
	12	347.2	76.1	456	23	US-09-543-407-29	Sequence 29, Appl
	13	315.2	69.1	361	6	US-08-233-642A-54	Sequence 54, Appl
	14	222.4	48.8	477	17	US-09-252-691-172	Sequence 172, App
	15	222.4	48.8	477	17	US-09-252-691C-172	Sequence 172, App
	16	222.4	48.8	477	51	US-10-417-886-172	Sequence 172, App
	17	216	47.4	456	23	US-09-543-407-3	Sequence 3, Appli
	18	216	47.4	675	13	US-08-978-878-3	Sequence 3, Appli
	19	216	47.4	675	31	US-09-741-873B-3	Sequence 3, Appli
	20	213.6	46.8	453	92	US-60-352-946-1	Sequence 1, Appli
	21	213.6	46.8	453	101	US-60-444-371-1	Sequence 1, Appli
	22	172	37.7	396	13	US-08-978-878-1	Sequence 1, Appli
	23	163.2	35.8	360	31	US-09-741-873B-1	Sequence 1, Appli
C	24	52.2	11.4	78	23	US-09-543-407-53	Sequence 53, Appl
	25	49.4	10.8	78	23	US-09-543-407-52	Sequence 52, Appl
C	26	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
	27	40.8	8.9	78	23	US-09-543-407-48	Sequence 48, Appl
	28	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
	29	40.6	8.9	78	23	US-09-543-407-49	Sequence 49, Appl
C	30	39.4	8.6	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
	31	39	8.6	100	45	US-10-146-492B-78	Sequence 78, Appl
	32	37.4	8.2	561	17	US-09-252-691-171	Sequence 171, App
	33	37.4	8.2	561	17	US-09-252-691C-171	Sequence 171, App
	34	37.4	8.2	561	51	US-10-417-886-171	Sequence 171, App
	35	37.4	8.2	1304	51	US-10-437-963-19773	Sequence 19773, A
	36	35.4	7.8	100	45	US-10-146-492B-77	Sequence 77, Appl
C	37	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 24330, A
C	38	35.4	7.8	3412	1	PCT-US01-08631-24948	Sequence 24948, A
	39	34.6	7.6	456	23	US-09-543-407-4	Sequence 4, Appli
C	40	34.2	7.5	385	18	US-09-333-535A-288	Sequence 288, App
C	41	34.2	7.5	385	25	US-09-615-606A-909	Sequence 909, App
C	42	34.2	7.5	396	64	US-60-089-524-4522	Sequence 4522, Ap
C	43	34.2	7.5	616	51	US-10-424-599-88945	Sequence 88945, A
C	44	34.2	7.5	880	27	US-09-654-617-2638	Sequence 2638, Ap
C	45	34.2	7.5	880	29	US-09-684-016-2638	Sequence 2638, Ap

ALIGNMENTS

```
RESULT 1
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 7.6e-136;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 2.6e-119;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |           | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 2.6e-119;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              |||||||||||||||||||| |           | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

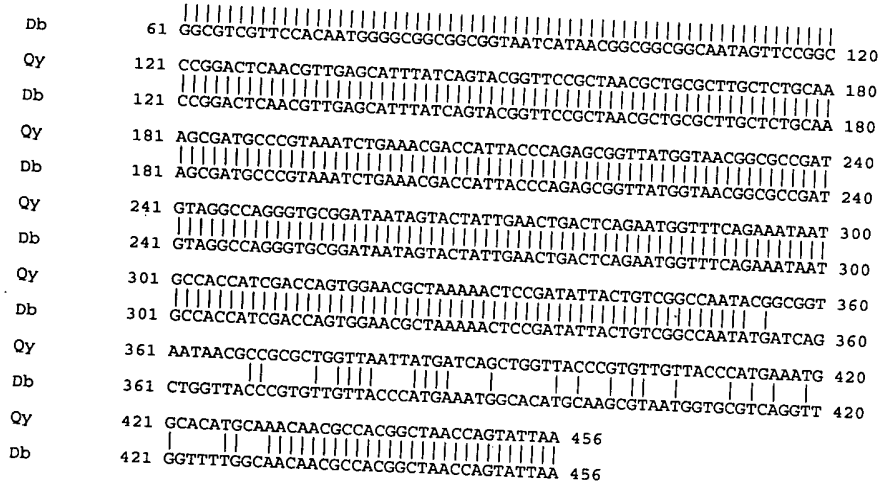
Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             82.5%;  Score 376;  DB 23;  Length 456;
  Best Local Similarity   89.0%;  Pred. No. 5.1e-110;
  Matches  406;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```



RESULT 5
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

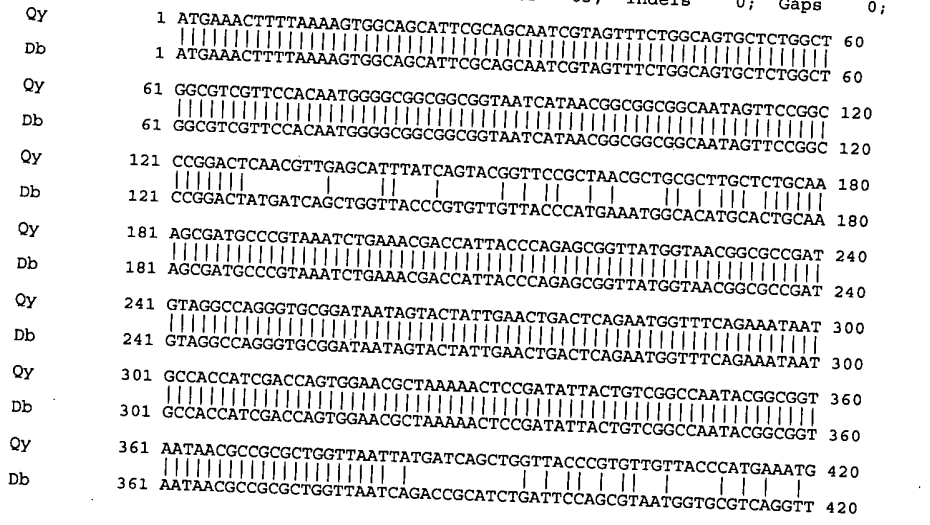
Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.3e-105;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;





RESULT 6
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.6e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;



```
Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 8.6e-103;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              || |    || |  |  ||| |          | | || | ||  |    | | |  |
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 8.6e-103;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||      |           ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||| |           | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.8e-102;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | |  |   |   ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||||| |           | | || | |  |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  |||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 9.3e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||| | |         | || |    || | |   | ||
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||           ||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||||||||||||||||| |           | | || | ||  |   | | |  |
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  |||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 9.3e-102;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 76.1%; Score 347.2; DB 23; Length 456;
Best Local Similarity 85.1%; Pred. No. 1e-100;
Matches 388; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 69.1%; Score 315.2; DB 6; Length 361;
Best Local Similarity 99.1%; Pred. No. 2e-90;
Matches 317; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATTA 383
              ||||||||||||||||||| |
Db        301 AACGCCGCGCTGGTTAATCA 320
```

RESULT 14
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-172

Query Match 48.8%; Score 222.4; DB 17; Length 477;
Best Local Similarity 68.0%; Pred. No. 1.9e-60;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| ||||||||||   | ||||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||      |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  |||||||||  ||||  | |||||||||  |||| | |||||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || ||  || || || ||  | ||| |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||   ||| |||| | ||| |  ||    || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||||  |||  |  |     |  |  |  ||  |  ||  | |  |  |  |
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  |||||||| || ||  |||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477
```

RESULT 15
US-09-252-691C-172

; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691C
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-172

Query Match 48.8%; Score 222.4; DB 17; Length 477;
Best Local Similarity 68.0%; Pred. No. 1.9e-60;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| ||||||||||   | ||||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||      |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  |||||||||  ||||  | |||||||||  |||| | |||||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || ||  || || || ||  | ||| |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||   ||| |||| | ||| |  ||    || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
              ||||  ||||  |||  |  |     |  |  |  ||  |  ||  | |  |  |  |
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  |||||||| || ||  |||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477
```

Search completed: March 19, 2004, 03:01:49
Job time : 2651.77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	216	47.4	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	172	37.7	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33.4	7.3	3990	5	US-09-830-230A-629	Sequence 629, App
	4	33.2	7.3	1355	6	US-10-045-674A-593	Sequence 593, App
	5	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
	6	31.4	6.9	1588	6	US-10-767-701-14522	Sequence 14522, A
c	7	31.2	6.8	290892	6	US-10-767-471-10809	Sequence 10809, A
c	8	30.2	6.6	512	6	US-10-781-014-257	Sequence 257, App
c	9	30.2	6.6	1326	6	US-10-781-014-255	Sequence 255, App
c	10	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
	11	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
c	12	29.4	6.4	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	13	29.4	6.4	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	14	29.4	6.4	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	15	29.4	6.4	1400	7	US-60-545-213-6407	Sequence 6407, Ap
	16	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	17	28.4	6.2	1350	6	US-10-784-480-291	Sequence 291, App
	18	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	19	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	20	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	21	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
	22	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
c	23	28	6.1	600	7	US-60-545-213-1313	Sequence 1313, Ap
c	24	28	6.1	600	7	US-60-545-213-1314	Sequence 1314, Ap
c	25	28	6.1	600	7	US-60-545-213-5585	Sequence 5585, Ap
c	26	28	6.1	600	7	US-60-545-213-5586	Sequence 5586, Ap
c	27	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
c	28	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
	29	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
	30	27.8	6.1	1580	6	US-10-451-467A-239	Sequence 239, App
c	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	32	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	33	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	34	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	35	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	36	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	37	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
c	38	27.4	6.0	1295	6	US-10-100-683-5480	Sequence 5480, Ap
	39	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
c	40	27.4	6.0	3416	6	US-10-100-683-3397	Sequence 3397, Ap
c	41	27.4	6.0	8206	6	US-10-021-698A-3529	Sequence 3529, Ap
	42	27.4	6.0	186449	6	US-10-021-698A-715	Sequence 715, App
	43	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
	44	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
	45	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 47.4%; Score 216; DB 5; Length 675;
Best Local Similarity 67.1%; Pred. No. 5.9e-62;
Matches 306; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442
Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502
Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 37.7%; Score 172; DB 5; Length 396;
Best Local Similarity 64.6%; Pred. No. 1.8e-47;
Matches 256; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
```

```
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300
Qy        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360
Qy        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
; LENGTH: 3990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1139)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1148)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1210)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1244)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1247)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1250)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1251)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

```
  Query Match             7.3%;  Score 33.4;  DB 5;  Length 3990;
  Best Local Similarity   47.4%;  Pred. No. 0.38;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  |||||| ||||     |    ||||      ||  ||| | | | |  | | |
Db       3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |    |    | ||||||||||  | |||     ||| ||
Db       3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |     | ||| |           | |||  |||||  |      |
Db       3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 3612

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
               || ||||       | |  ||| | || |
Db       3613 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 3643


RESULT 4
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674A
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 06/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674A-593

  Query Match             7.3%;  Score 33.2;  DB 6;  Length 1355;
  Best Local Similarity   23.9%;  Pred. No. 0.29;
  Matches   83;  Conservative   91;  Mismatches  174;  Indels    0;  Gaps    0;

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137
              :||:||:||: ::  :  :  : ::||:||:  ::::   ||:  | :::   :    |
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATC 197
              :  :: :  :  :||::::| :: :| :|   |:| :    |:|     ||   :
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        198 TGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGA 257
              : |:  : |:|| ||: |:| :| : |:|  ||:|      |    | :      |   |
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        258 TAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              :  :::: | |       : |   || | :|   |:  :  :| :  :||:   ::
Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYATYGGYGAYGTYWS 1007

Qy        318 GAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGT 377
              :  :   ||:  : |:  : |:  :||  | | :|  ||:  :||  |:  : |||
Db       1008 YGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKAATTCYCARATGGC 1067

Qy        378 TAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 425
              : |   :| : |   :| :| : ::  : |:|   |: |  |   :||
Db       1068 YCARGTYGGWGAYGGKGAYAAYWSWCCKYTWATGAAYAAYTTYMGWCA 1115


RESULT 5
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
;  APPLICANT: Neubauer, Antonie
;  APPLICANT:  Ziegler, Christina
;  TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
;  FILE REFERENCE: 1/1372
;  CURRENT APPLICATION NUMBER: US/10/624,149A
;  CURRENT FILING DATE:  2003-07-21
;  PRIOR APPLICATION NUMBER: 60/403,282
;  PRIOR FILING DATE: 2002-08-14
;  PRIOR APPLICATION NUMBER: DE 10233064
;  PRIOR FILING DATE: 2002-07-19
;  PRIOR APPLICATION NUMBER: DE 10317008
;  PRIOR FILING DATE: 2003-04-11
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
;   LENGTH: 150223
;   TYPE: DNA
;   ORGANISM: Equine herpesvirus 1
;  PUBLICATION INFORMATION:
;   AUTHORS: Telford,E.A.
;   AUTHORS: Watson, M.S.
;   AUTHORS: McBride, K.
;   AUTHORS: Davison, A.J.
;   TITLE: The DNA sequence of equine herpesvirus-1
;   JOURNAL: Virology
;   VOLUME: 189
;   ISSUE: 1
;   PAGES: 304-316
;   DATE: JUL-1992
;   DATABASE ACCESSION NUMBER: NC 001491, NCBI
;   DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 150223;
  Best Local Similarity   51.4%;  Pred. No. 3.6;
  Matches   75;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              || ||  ||  |  | | || ||  | ||| ||| |    |  ||| |||   | ||  |||
Db      95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy        233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
              |       |  |||| ||  ||||||    | || |  | | | ||| |    |  |||
Db      95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy        293 GAAATAATGCCACCATCGACCAGTGG 318
              || |   |      || ||| | || |
Db      95662 GAGAGCTTATTCACAACGAGCTGTTG 95687


RESULT 6
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
```

```
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
;   LENGTH: 1588
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 1588;
  Best Local Similarity   44.5%;  Pred. No. 1.2;
  Matches  125;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
Db        455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
Db        515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy        160 AACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
Db        575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
Db        635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAA 320
Db        695 GCTGGTAGCTTTGGCGGTGGTTCCACTGGCGCCGAGAAGGA 735


RESULT 7
US-10-767-471-10809/c
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
;   LENGTH: 290892
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(290892)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 290892;
  Best Local Similarity   44.6%;  Pred. No. 12;
  Matches  123;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
Db     234133 GATTATGATGATGATGATGGTAATGGTGATGATGGTGATGGTGATAGTTACAGTAGTGAT 234074

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
Db     234073 GTTGGTAATGATAGTGGTGGTGACGACGGTGATGCTAATGATTATGATGATGATGGTGGT 234014

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
Db     234013 GGTAATGGTGATGATGGTAATGTTGGTGATGATCGTGGCTGTGATGGTGATGGTTATGAT 233954

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGGTT 396
Db     233953 GGGAATGTTGGTGAAGATCGTGGCTGTGATGGTGATGGTGGTGATGATGATGACAATGAT 233894

Qy        397 ACCCGTGTTGTTACCCATGAAATGGCACATGCAAAC 432
Db     233893 AAGGATGGTAATGAAAATGATAATGCTGATGCTAAC 233858


RESULT 8
US-10-781-014-257/c
; Sequence 257, Application US/10781014
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
;  TITLE OF INVENTION:  INVOLVED IN CARBON METABOLISM AND ENERGY
;  TITLE OF INVENTION:  PRODUCTION
;  FILE REFERENCE: BGI-126CPCN
;  CURRENT APPLICATION NUMBER: US/10/781,014
;  CURRENT FILING DATE:  2004-02-17
;  PRIOR APPLICATION NUMBER: US 09/602,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/143,208
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: 60/151,572
;  PRIOR FILING DATE: 1999-08-31
;  PRIOR APPLICATION NUMBER: DE 19931412.8
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931413.6
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931424.1
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931428.4
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 257
;   LENGTH: 512
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(489)
;   OTHER INFORMATION: FRXA02596
US-10-781-014-257

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 512;
  Best Local Similarity   56.6%;  Pred. No. 1.9;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        310 GACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
Db        375 GCCCAGTCGACCGCCGAACAGCACCTTATTATTAGCAGCCTCTTCAGCAGCCAGAAGGCG 316

Qy        370 GCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTT 408
Db        315 GTACTGCTTCAGCATGTCTCGGTCGTCTGGAGTGTTGAT 277
```

```
RESULT 9
US-10-781-014-255/c
; Sequence 255, Application US/10781014
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
;  TITLE OF INVENTION:  INVOLVED IN CARBON METABOLISM AND ENERGY
;  TITLE OF INVENTION:  PRODUCTION
;  FILE REFERENCE: BGI-126CPCN
;  CURRENT APPLICATION NUMBER: US/10/781,014
;  CURRENT FILING DATE:  2004-02-17
;  PRIOR APPLICATION NUMBER: US 09/602,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/143,208
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: 60/151,572
;  PRIOR FILING DATE: 1999-08-31
;  PRIOR APPLICATION NUMBER: DE 19931412.8
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931413.6
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931424.1
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931428.4
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 255
;   LENGTH: 1326
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101)..(1303)
;   OTHER INFORMATION: RXN02596
US-10-781-014-255

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 1326;
  Best Local Similarity   56.6%;  Pred. No. 2.8;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        310 GACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              | ||||| || ||| || |   |  ||||| |  | |||  | | || |  | || ||
Db       1189 GCCCAGTCGACCGCCGAACAGCACCTTATTATTAGCAGCCTCTTCAGCAGCCAGAAGGCG 1130

Qy        370 GCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTT 408
              |  ||| || |  |||   .| | ||     ||||||
Db       1129 GTACTGCTTCAGCATGTCTCGGTCGTCTGGAGTGTTGAT 1091


RESULT 10
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
;   LENGTH: 24900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(24900)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 24900;
  Best Local Similarity   52.4%;  Pred. No. 15;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||  | || ||  | || | ||  |       |   | || || |  | | ||||| |
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |  |||||  |  | | |||||   | ||||| ||| | | | |     || |    |||
Db       9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929

Qy        327 AAAC 330
              || |
Db       9928 AATC 9925


RESULT 11
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 2.8;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||     |       |||| | |     | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
               |||||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 12
US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
```

```
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2134
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2134

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 5.3;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGC 391
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAG 450
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 13
US-60-545-213-2135/c
; Sequence 2135, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2135
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-2135

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 5.3;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGC 391
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAG 450
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 14
US-60-545-213-6406/c
; Sequence 6406, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6406
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6406

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 5.3;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGC 391
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAG 450
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 15
US-60-545-213-6407/c
; Sequence 6407, Application US/60545213
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William Martin
;  TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
;  TITLE OF INVENTION:  Target Genes
;  FILE REFERENCE: AM101083 (031896-042099)
;  CURRENT APPLICATION NUMBER: US/60/545,213
;  CURRENT FILING DATE:  2004-02-18
;  NUMBER OF SEQ ID NOS: 303284
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6407
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-60-545-213-6407

  Query Match             6.4%;  Score 29.4;  DB 7;  Length 1400;
  Best Local Similarity   45.2%;  Pred. No. 5.3;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db        437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGC 391
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAG 450
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199
```

Search completed: March 19, 2004, 03:10:07
Job time : 55.1445 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-11
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0% .
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	216.6	47.5	680	12	BJ618688	BJ618688 BJ618688
	2	35.2	7.7	813	28	BZ204853	BZ204853 CH230-390
	3	34.8	7.6	801	12	BI648266	BI648266 603278209
	4	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
	5	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	6	34	7.5	234	10	BB581987	BB581987 BB581987
c	7	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	8	34	7.5	811	28	BZ469000	BZ469000 BOOAO58TF
	9	33.8	7.4	296	9	AV058630	AV058630 AV058630
c	10	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	11	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	12	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
c	13	33.8	7.4	695	28	AQ637195	AQ637195 RPCI-11-4
	14	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	15	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	16	33.6	7.4	374	13	BY236219	BY236219 BY236219
	17	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	18	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
c	19	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	20	33.4	7.3	505	6	AL809250	Al809250 Triticum
	21	33.4	7.3	624	14	CB578319	CB578319 AMGNNUC:N
	22	33.4	7.3	664	13	BX506928	BX506928 DKFZp779I
	23	33.4	7.3	679	29	CC961526	CC961526 BOIGS10TF
c	24	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
c	25	33.4	7.3	699	12	BI959328	BI959328 HVSMEn001
c	26	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF
c	27	33.4	7.3	709	28	BH690914	BH690914 BOHXR64TF
	28	33.4	7.3	715	13	BU444683	BU444683 603765715
c	29	33.4	7.3	785	28	BH578793	BH578793 BOGLY20TR
c	30	33.4	7.3	815	28	BH461716	BH461716 BOHIV03TR
c	31	33.4	7.3	827	28	BH438448	BH438448 BOGVQ28TR
c	32	33.4	7.3	830	28	BH715216	BH715216 BOMIM42TR
c	33	33.4	7.3	835	28	BH575194	BH575194 BOHJO27TF
c	34	33.4	7.3	853	28	BH578402	BH578402 BOHNV61TF
	35	33.4	7.3	856	14	CD456845	CD456845 Fg05_03n1
c	36	33.4	7.3	1026	28	BZ458784	BZ458784 BONHB14TF
	37	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
	38	33.2	7.3	592	29	CG072975	CG072975 PUICN43TB
	39	33.2	7.3	629	9	AU295325	AU295325 AU295325
	40	33.2	7.3	675	28	AQ361497	AQ361497 mgxb0004K
	41	33.2	7.3	788	9	AU139321	AU139321 AU139321
c	42	33.2	7.3	1646	10	BF579120	BF579120 602096054
	43	33	7.2	456	12	BJ331339	BJ331339 BJ331339
	44	33	7.2	524	12	BJ367372	BJ367372 BJ367372
	45	33	7.2	547	12	BJ333958	BJ333958 BJ333958

ALIGNMENTS

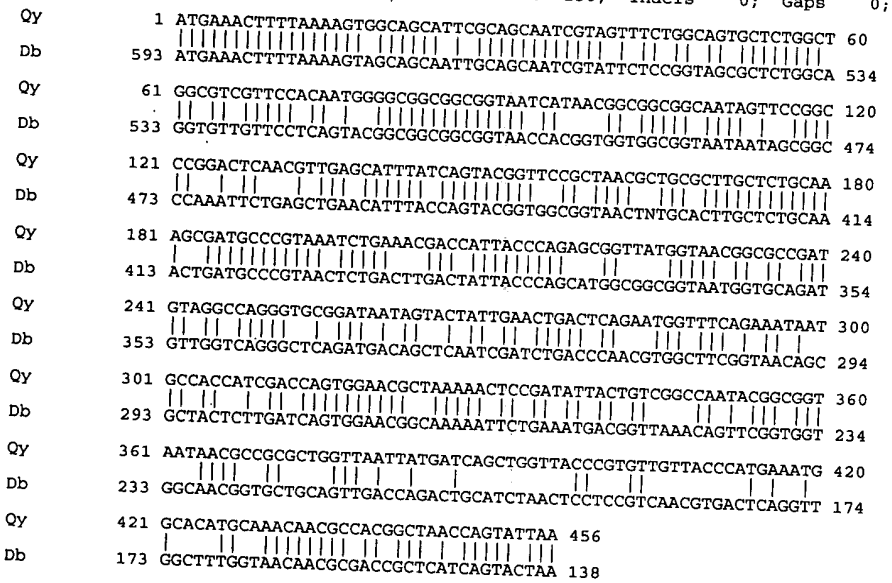
RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.

FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 47.5%; Score 216.6; DB 12; Length 680;
Best Local Similarity 67.1%; Pred. No. 8e-51;
Matches 306; Conservative 0; Mismatches 150; Indels 0; Gaps 0;



RESULT 2
BZ204853
LOCUS BZ204853 813 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-390D21, genomic survey sequence.
ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 813)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-390D21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 390 row: D column: 21
Seq primer: T7
Class: BAC ends.

FEATURES Location/Qualifiers
source 1. .813
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"

ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 813;
Best Local Similarity 49.0%; Pred. No. 59;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

RESULT 3
BI648266
LOCUS BI648266 801 bp mRNA linear EST 12-SEP-2001
DEFINITION 6032782O9F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5318307 5', mRNA sequence.
ACCESSION BI648266
VERSION BI648266.1 GI:15562502
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 801)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

```
            http://image.llnl.gov
            Plate: LLAM11807  row: b  column: 04
            High quality sequence stop: 798.
FEATURES              Location/Qualifiers
     source           1. .801
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="129,C57BL/6J,FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5318307"
                      /tissue_type="tumor, gross tissue"
                      /dev_stage="10 months"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_Mam3"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigators
                      providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                      Reference for transgenic model: Xu et al., Nature Genetics
                      22, 37-43 (1999)."
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 12;  Length 801;
  Best Local Similarity   50.0%;  Pred. No. 76;
  Matches  114;  Conservative    0;  Mismatches  112;  Indels    2;  Gaps    1;

Qy        210 TACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTAT 269
              | | |  |||    ||||| || ||  |         | ||  | ||  ||  | ||
Db        322 TTCACCCAGCAAGCATGGTGACCGCTTCATCCCCTCGCGGGCCGGGGCCAACTGGACGCG 381

Qy        270 TGAACTGACTCAGAATGGTTTCAGAA--ATAATGCCACCATCGACCAGTGGAACGCTAAA 327
              |||||| | ||| ||   | || | ||  || ||| | |  |  | || || ||
Db        382 TGAACTTCCACAGGATCAATGAACAATGAGAAGTCCCCCAGCCAGAACCGCAAAGCCAAG 441

Qy        328 AACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGAT 387
               || ||          ||| || |||||       || |||| ||| | ||  ||||
Db        442 GACGCCACCTCGGACAATGGCAAAGACGGCCTGGCCTACTCCGCACTGCTGAAGAATGAG 501

Qy        388 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAAC 435
              | ||||| | || |  | |  |    | |     ||||  |  |  ||
Db        502 CTGCTGGGTGCCGGCATCGAGAAGGTTCAGGACCCACAGACGGAGGAC 549


RESULT 4
BJ334624
LOCUS       BJ334624                523 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda47o06 5', mRNA sequence.
ACCESSION   BJ334624
VERSION     BJ334624.1  GI:19164754
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .523
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="dda47o06"
                      /sex="mat A"
                      /dev_stage="Aggregation stage"
                      /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 523;
  Best Local Similarity   57.1%;  Pred. No. 93;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        398 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGANGACAATAAT 457

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACT 342
              |||  || || | || |    ||   ||||||     |||||| |
Db        458 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAAT 502


RESULT 5
BG671547
LOCUS       BG671547                1036 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5',
            mRNA sequence.
ACCESSION   BG671547
VERSION     BG671547.1  GI:13893646
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C.,
            Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D.,
            Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
   PUBMED   12060780
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .1036
                      /organism="Rattus norvegicus"
                      /mol_type="mRNA"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="DRNBUB05"
                      /sex="male"
                      /tissue_type="dorsal root ganglion"
                      /dev_stage="adult"
                      /clone_lib="Rat DRG Library"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 1036;
  Best Local Similarity   49.2%;  Pred. No. 1.3e+02;
  Matches   90;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;
```

```
Qy          11 TAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTC 70
               | ||||   ||  | |  |||    |||  || | | || |||| |      ||||| |
Db         168 TGAAAGATACATGAGTGACAGTGTCCGTGATTGCAGACATTGCTGTATGATGCGTCCTGT 227

Qy          71 CACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAA 130
               |   | | | |  ||  ||||| ||  ||   | |  || | |  |    ||     |
Db         228 CTGTAGGAGTCCTCGCTGGTAACGATGCCGAGAGTGCACATCGATATCGGTGGGCGGTTA 287

Qy         131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
               | |     |  |  ||  | |    | ||| |||   ||| ||||  ||| || ||
Db         288 CATCTGAAGTGGAAGAGGTGGCTCTTTCGAACTCTGACTTTGATCTGTCAAGGGACGCTT 347

Qy         191 GTA 193
               |||
Db         348 GTA 350
```

RESULT 6
BB581987/c
LOCUS BB581987 234 bp mRNA linear EST 30-NOV-2000
DEFINITION BB581987 RIKEN full-length enriched, adult male colon Mus musculus cDNA clone 9030005O05 5', mRNA sequence.
ACCESSION BB581987
VERSION BB581987.1 GI:11478531
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 234)
AUTHORS Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .234
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9030005O05"
/sex="male"
/tissue_type="colon"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male colon"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match 7.5%; Score 34; DB 10; Length 234;
Best Local Similarity 51.3%; Pred. No. 73;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy          35 TCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATC 94
               | ||  ||  | |    | |   || ||  || |    ||  || | || || ||    |
Db         200 TAGTTTTTGGTAGGTCGGTTTATGCGGGTTTCTTCTTCCACCAGGCCCGCCGCCGTTTCC 141

Qy          95 ATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTT 154
                ||||  |||  ||  ||   ||         |   |  || | |||||||  ||  ||  ||
Db         140 CGAACGCCGGGTGCTTTATAGCCTTGTTCAAAATATGGTAGCGCATTTACCACCACCTTT 81

Qy         155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
               |||  | ||   || | ||| |  ||  |||  ||
Db          80 CCGGTTACCAAGCCCGTGCCCGCCATTGCGCGGC 47
```

RESULT 7
CA711477/c
LOCUS CA711477 500 bp mRNA linear EST 26-NOV-2002
DEFINITION wdk2c.pk014.o3 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.o3 5' end, mRNA sequence.
ACCESSION CA711477
VERSION CA711477.1 GI:25433270
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 500)
AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE DuPont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.

FEATURES Location/Qualifiers
source 1. .500
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"

```
                /clone="wdk2c.pk014.o3"
                /tissue_type="kernel"
                /clone_lib="wdk2c"
                /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
                days after anthesis."
ORIGIN

  Query Match             7.5%;  Score 34;  DB 14;  Length 500;
  Best Local Similarity   54.5%;  Pred. No. 1e+02;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  |  ||| |||||||   ||   |  | ||| ||       | | ||  | ||| |
Db        274 AGNTGNTGCAGTCGCAGCCNCCGGTCTCCCCGGCCGTNTCGAACCCGCCGGAGCTCCCCN 215

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGAC 126
                 ||| ||||| ||  || |  | |||||||||     |    ||||   ||
Db        214 TGTGGGAGGCGGTGGGCATGACCATGGCGGCGGCGCCGGCGGTGGCCNCCAC 163


RESULT 8
BZ469000/c
LOCUS       BZ469000                811 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BOOAO58TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOAO58,
            genomic survey sequence.
ACCESSION   BZ469000
VERSION     BZ469000.1  GI:26764546
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOOAO58TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOOAO58"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 28;  Length 811;
  Best Local Similarity   48.9%;  Pred. No. 1.3e+02;
  Matches   91;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        176 TGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCG 235
              ||  | || |     ||||| |||       ||||| |     |  | |||||  ||  ||
Db        610 TGTTAGGCAAGCATTGTAAAACTGTGTTTCCCATTTCATCTACTGTTTATGACAAATGCT 551

Qy        236 CCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
                ||| ||   ||   ||  | |   | | ||  |    || ||  | || ||||  | |
Db        550 TTGATCTAATTCACTCTGATGTTTGGACTGCTCCTTGCTTGTCTAGGGATAGTTTTAAAT 491

Qy        296 ATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 355
                |  || |||  ||    |   ||  | ||| || ||   || ||  |    |||  |
Db        490 ACTTTGTCACATTCATAAATGAAAAATCGAAATACACCTGGATAACACTTATTCAAACCA 431

Qy        356 GCGGTA 361
                 | ||
Db        430 AAGATA 425


RESULT 9
AV058630
LOCUS       AV058630                296 bp    mRNA    linear   EST 23-JUN-1999
DEFINITION  AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA
            clone 1810054H02, mRNA sequence.
ACCESSION   AV058630
VERSION     AV058630.1  GI:5158377
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 296)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .296
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1810054H02"
                     /sex="male"
                     /tissue_type="pancreas"
                     /dev_stage="adult"
                     /clone_lib="Mus musculus pancreas C57BL/6J adult"
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 296;
  Best Local Similarity   53.4%;  Pred. No. 93;
  Matches   71;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy        251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCG 310
              |||  ||  | |||  | |  |  | ||  ||  ||| |  |||||  | ||  | |  |
Db         13 GTGGAGACCACAGTTTTTTCTAGGTCACCAAGGCTGGGTAAAGAAAAGAAGCATCAACTG 72

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              ||||||  | | | |||||     ||     | ||     || | | |  || | ||| |  |
Db         73 TCCAGTGAAGCACAAAAAATGAAGAAGAAAATGATTACCCAAAAGTTGGAATTAAGGTAG 132

Qy        371 CGCTGGTTAATTA 383
                | ||  || ||
Db        133 AGATGCCAAAATA 145
```

```
RESULT 10
AA746477/c
LOCUS       AA746477                 365 bp    mRNA    linear   EST 27-JAN-1998
DEFINITION  nw62e02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
            similar to TR:O00555 O00555 NEURONAL CALCIUM CHANNEL ALPHA 1A
            SUBUNIT ISOFORM A-1. ;contains element TAR1 repetitive element ;,
            mRNA sequence.
ACCESSION   AA746477
VERSION     AA746477.1  GI:2786463
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 661    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1251194"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 365;
  Best Local Similarity   56.9%;  Pred. No. 1e+02;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  ||||||   |||||||   | ||   | ||| ||| | |||| |||   |  ||||
Db        119 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGTGGCCAGGCCGGCCGGGCGCCACC 60

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |  || |  ||| ||||  ||   | ||   | | |   | | |||| |
Db         59 AGCGGCCCTCGGAGGTACCCAGGCCCGCACGGCCGACGCTACTGGCCGG 11

RESULT 11
```

```
CD374421/c
LOCUS       CD374421                 558 bp    mRNA    linear   EST 19-AUG-2003
DEFINITION  TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)
            Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION   CD374421
VERSION     CD374421.1  GI:31229939
KEYWORDS    EST.
SOURCE      Wuchereria bancrofti
  ORGANISM  Wuchereria bancrofti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Wuchereria.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
  TITLE     Toward the characterization of the Wuchereria bancrofti
            microfilarial transcriptome with comparisons to those of Brugia
            malayi and Onchocerca volvulus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Thomas B. Nutman
            Laboratory of Parasitic Diseases
            NIH
            Building 4 Room 126, Bethesda, MD 20892-0425, USA
            Email: tnutman@niaid.nih.gov
            Seq primer: pBluescript T3.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Wuchereria bancrofti"
                     /mol_type="mRNA"
                     /db_xref="taxon:6293"
                     /clone="TNWbmfC2H3"
                     /dev_stage="microfilaria"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="Wuchereria bancrofti microfilaria cDNA
                     (SAW95SjL-WbMf)"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 85,000 microfilariae
                     isolated from the blood of an infected individual from
                     Guyana, South America and converted to double-stranded
                     cDNA using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNA pol I. The library has 2.9 x 10E5
                     independent recombinants and the average insert size is
                     ~1kb. The library was constructed by Sandra J. Laney. The
                     library is available from Dr.S.A.Williams, email:
                     genome@smith.edu."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 558;
  Best Local Similarity   60.7%;  Pred. No. 1.2e+02;
  Matches   65;  Conservative    0;  Mismatches   40;  Indels    2;  Gaps    1;

Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              ||| |||| |||||||| |     || ||||| |||| |    | ||  | |
Db        161 TGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCCCAAATTCTANGCT 102

Qy        136 AGCATTTATC--AGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||| |  ||  ||||  ||  | | |||| ||||||||| ||
Db        101 GACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55

RESULT 12
BU765156/c
LOCUS       BU765156                 559 bp    mRNA    linear   EST 10-OCT-2002
DEFINITION  sas23f11.y1 Gm-c1080 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1080-4869 5' similar to SW:D100_ARATH Q00874
            DNA-DAMAGE-REPAIR/TOLERATION PROTEIN DRT100 PRECURSOR ;, mRNA
            sequence.
ACCESSION   BU765156
VERSION     BU765156.1  GI:23733884
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
```

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 559)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: ccu@resgen.com web site: www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 448.
FEATURES Location/Qualifiers
source 1. .559
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1080-4869"
/tissue_type="Roots of 8 day old 'Bragg' supernodulating mutant NTS382 seedlings"
/dev_stage="8 days old"
/lab_host="DH10B"
/clone_lib="Gm-c1080"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The mRNA was isolated from roots of 8 day old 'Bragg' supernodulating mutant NTS382 seedlings that were infected with Bradyhizobium japonicum, strain USDA 110, 72 hours prior to harvest. Dr. Gary Stacey generously donated the tissue. The roots were flash-frozen in liquid nitrogen. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An 'anchor' nucleotide (V=A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first- strand synthesis primer was then restricted by digestion with XhoI from Promega (40U/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using Sephacryl S-500 High Resolution (Pharmacia Biotech) in a 2-mm diameter column and a bed volume of approximately 1ml. The column eluent was precipitated, redissolved, and ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). This library was constucted in the laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at Northern Arizona University."
ORIGIN

Query Match 7.4%; Score 33.8; DB 13; Length 559;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy        76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
             ||||| || | |    |||| | ||  ||||||| ||| | |  || |||     ||||
Db       233 TGGGGTCGCAGGCGACGCCATACCAGCTCCGGCAACAGTACAGTGCCCGACCATGTGTTG 174

Qy       136 AGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGT 192
             |  ||     |||| |||||    |  || || |   | || | |||||||
Db       173 AAGATGCCCAAGTAGGGTTCGATCAGGGCAGCTTTGAAGGCCAAGAGTGCTGCCCGT 117
```

RESULT 13
AQ637195/c
LOCUS AQ637195 695 bp DNA linear GSS 17-JUN-1999
DEFINITION RPCI-11-481M9.TV RPCI-11 Homo sapiens genomic clone RPCI-11-481M9, genomic survey sequence.
ACCESSION AQ637195
VERSION AQ637195.1 GI:5099830
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 695)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-481M9.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .695
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7684616"
/db_xref="taxon:9606"
/clone="RPCI-11-481M9"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
ORIGIN

Query Match 7.4%; Score 33.8; DB 28; Length 695;
Best Local Similarity 49.2%; Pred. No. 1.4e+02;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```
Qy       275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
             | ||| | | ||      || | |    ||   | ||||  ||  |  | || |
Db       291 TAACTAATACTTGTACACACAAAAGAGGATTAATGATCAAGTGTCACATTCTAGACACGT 232

Qy       335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATTATGATCAGCTGG 394
             || |||| || |       | |    ||   | ||   ||||   | | |   |||    |
Db       231 ATTTTACAGTTGTAACTCAAGTGCATACAGATGCACAGCTGTACACTCACACTCATACAG 172

Qy       395 TTACCCGTGTTGTTACCCATGAAATGGCACATGCAAACAACGCCACGGCTAACCAGTATT 454
             |  | ||| |  || ||| ||| | |||||  || ||  | |||   ||  || | |
Db       171 CCATACATGTGGACACACATTAAAAGTCACATATAAGCATGGACACAAGTATACATTCAT 112
```

```
Qy        455 A 455
              |
Db        111 A 111
```

```
RESULT 14
CG178280
LOCUS       CG178280                941 bp    DNA     linear   GSS 21-AUG-2003
DEFINITION  PUJDF79TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0641N14,
            genomic survey sequence.
ACCESSION   CG178280
VERSION     CG178280.1  GI:34069341
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 941)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUJDF79TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .941
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa0641N14"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN
```

```
  Query Match             7.4%;  Score 33.8;  DB 29;  Length 941;
  Best Local Similarity   51.0%;  Pred. No. 1.6e+02;
  Matches   80;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | ||||  | | || || | || || | || || || | ||  ||    | | ||||| |
Db        236 TTGCAGTCTACACATCAGTGGTCGTATATGCCACAGTTCCTGCACCGGCCTTTCCGCCGC 295

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGA 136
              || |   | | || |||  ||| | || |   |   | ||| |||||  ||   ||
Db        296 GGAGTACCCTCCGTCGTCACCACGACCGCCGGTCTTCGACTTGCCGGGACTGTACCCCGA 355

Qy        137 GCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGC 173
                 || ||   || | ||  |    |  | || ||
Db        356 CTTCTTCTCGCCACCGCTCAACGGCGACGACTCTGGC 392
```

```
RESULT 15
CNS02GQP/c
LOCUS       CNS02GQP                1098 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            138K23 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL196666
VERSION     AL196666.1  GI:7834816
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,P.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1098)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1098
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="138K23"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG138AF12SP1~end :
                     PUC-Ori"
ORIGIN
```

```
  Query Match             7.4%;  Score 33.8;  DB 29;  Length 1098;
  Best Local Similarity   55.8%;  Pred. No. 1.7e+02;
  Matches   53;  Conservative    5;  Mismatches   37;  Indels    0;  Gaps    0;

Qy         19 GCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGG 78
              | ||   | || | |   | | || ||||    | |||||||   | |||| |:|
Db        773 GAAGTCCTGGCTGAACAGGGGGATTTTGGCGTAACGCTGGCTGAAATGGTTCAAMACRAT 714

Qy         79 GGCGGCGGCGGTAATCATAACGGCGGCGGCAATAG 113
              :  ||||||| |:||| |:||| |: || | || |
Db        713 RAAGGCGGCGTTMATCCTMACGCCRACGCCGATGG 679
```

```
Search completed: March 18, 2004, 20:11:25
Job time : 2034.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		404.8	88.8	456	6	I44909	I44909 Sequence 58
2		404.8	88.8	2067	1	SEU43280	U43280 Salmonella
3		396.8	87.0	5103	1	STAJ2301	AJ002301 Salmonell
4		396.8	87.0	22411	1	AE008749	AE008749 Salmonell
5		395.2	86.7	254050	1	AL627269	AL627269 Salmonell
6	c	395.2	86.7	301983	1	AE016840	AE016840 Salmonell
7		389.4	85.4	1048	1	STAGFBA	AJ000514 Salmonell
8		287.2	63.0	361	6	I44908	I44908 Sequence 56
9		281.6	61.8	2889	1	CSP515700	AJ515700 Citrobact
10		229.2	50.3	2920	1	CFR515701	AJ515701 Citrobact
11		229	50.2	230	1	SEU53207	U53207 Salmonella
12		227.2	49.8	4680	1	ECCSGABDG	X90754 E.coli csgG
13		227.2	49.8	10346	1	AE000205	AE000205 Escherich
14		227.2	49.8	15047	1	D90741	D90741 Escherichia
15		225.6	49.5	456	6	AX814811	AX814811 Sequence
16		225.6	49.5	648	1	ECOCSGAA	L04979 Escherichia
17		220.6	48.4	306358	1	AE016759	AE016759 Escherich
18		217.4	47.7	1711	1	AF275733	AF275733 Escherich
19		217.4	47.7	10190	1	AE005315	AE005315 Escherich
20		217.4	47.7	327773	1	AP002554	AP002554 Escherich
21		188.8	41.4	2883	1	ESA515702	AJ515702 Enterobac
22		163.2	35.8	10370	1	AE015131	AE015131 Shigella
23		163.2	35.8	292504	1	AE016981	AE016981 Shigella
24	c	140.2	30.7	437	1	AF237726	AF237726 Shigella
25		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
26		64.4	14.1	19201	1	D90742	D90742 Escherichia
27		44.8	9.8	72	6	AX814798	AX814798 Sequence
28	c	44.8	9.8	2000	6	AX655393	AX655393 Sequence
29		39.4	8.6	956	8	AK058943	AK058943 Oryza sat
30		39.4	8.6	2720	8	AK102629	AK102629 Oryza sat
31		39.4	8.6	152448	8	AP004339	AP004339 Oryza sat
32		38.8	8.5	13264	1	AE001012	AE001012 Archaeogl
33		38.2	8.4	125020	9	AF429315	AF429315 Homo sapi
34	c	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
35	c	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
36	c	36	7.9	168267	2	AC146009	AC146009 Pan trogl
37		35.8	7.9	172600	9	AL359644	AL359644 Human DNA
38	c	35.8	7.9	193168	2	AC011969	AC011969 Homo sapi
39		35.6	7.8	301214	1	AE016786	AE016786 Pseudomon
40	c	35	7.7	117541	2	AC104285	AC104285 Oryza sat
41		34.4	7.5	560	8	AY247273	AY247273 Medicago
42		34.4	7.5	1553	1	HAARR16S	M59123 Haloanaerob
43		34.4	7.5	2751	6	AX416572	AX416572 Sequence
44		34.4	7.5	151828	2	AC146008	AC146008 Pan trogl
45	c	34.4	7.5	160449	10	AC130549	AC130549 Mus muscu

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match              88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity    93.0%;  Pred. No. 8.9e-106;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||     | ||||  ||         |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match              88.8%;  Score 404.8;  DB 1;  Length 2067;
  Best Local Similarity    93.0%;  Pred. No. 1e-105;
```

```
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||        |  || | ||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVPYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
     gene            3344. .3376
                     /gene="csgBA"
     -35_signal      3344. .3349
                     /gene="csgBA"
     -10_signal      3371. .3376
                     /gene="csgBA"
     gene            3469. .3924
                     /gene="csgB"
     CDS             3469. .3924
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="CAA05316.1"
                     /db_xref="GI:2739238"
                     /db_xref="GOA:P55226"
                     /db_xref="SWISS-PROT:P55226"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     3469. .3531
                     /gene="csgB"
     mat_peptide     3532. .3921
                     /gene="csgB"
                     /product="CsgB protein"
                     /function="nucleator"
     gene            3966. .4421
                     /gene="csgA"
     CDS             3966. .4421
                     /gene="csgA"
                     /function="fiber subunit"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="CAA05317.1"
                     /db_xref="GI:2739239"
                     /db_xref="GOA:P55225"
                     /db_xref="SWISS-PROT:P55225"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     3966. .4025
                     /gene="csgA"
     mat_peptide     4026. .4418
                     /gene="csgA"
                     /product="CsgA protein"
                     /function="fiber subunit"
     gene            4483. .4809
                     /gene="csgC"
     CDS             4483. .4809
                     /gene="csgC"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgC protein"
                     /protein_id="CAA05318.1"
                     /db_xref="GI:2739240"
                     /db_xref="SWISS-PROT:P55227"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
     sig_peptide     4483. .4533
                     /gene="csgC"
     mat_peptide     4534. .4806
                     /gene="csgC"
                     /product="CsgC"
     gene            4940. .5103
     CDS             /gene="ORF103"
                     4940. .>5103
                     /gene="ORF103"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA05319.1"
                     /db_xref="GI:2739241"
                     /db_xref="SWISS-PROT:O54295"
                     /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                     CDISTHAENI"
     ORIGIN
```

```
  Query Match             87.0%;  Score 396.8;  DB 1;  Length 5103;
  Best Local Similarity   91.9%;  Pred. No. 2.3e-103;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||| |||||  |
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||          |  || | ||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421
```

```
RESULT 4
AE008749
LOCUS       AE008749                22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
     RBS                complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene               complement(2580. .4088)
                        /gene="STM1128"
     CDS                complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS                complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene               complement(4413. .5099)
                        /gene="STM1129"
     CDS                complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
     RBS                complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene               5599. .6759
                        /gene="STM1130"
     CDS                5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene               6793. .7497
                        /gene="STM1131"
     RBS                6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match             87.0%;  Score 396.8;  DB 1;  Length 22411;
  Best Local Similarity   91.9%;  Pred. No. 2.6e-103;
  Matches  419;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||        |  || | ||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
  PUBMED    11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAININSSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDPELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             86.7%;  Score 395.2;  DB 1;  Length 254050;
  Best Local Similarity   91.7%;  Pred. No. 9.6e-103;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||       |  || | ||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361
```

RESULT 6
AE016840/c
LOCUS AE016840 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of 16 of the complete genome.
ACCESSION AE016840 AE014613
VERSION AE016840.1 GI:29137797
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi Ty2
ORGANISM Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 301983)
AUTHORS Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and CT18
JOURNAL J. Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE 22531367
PUBMED 12644504
REFERENCE 2 (bases 1 to 301983)
AUTHORS Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES Location/Qualifiers

```
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
     CDS             /locus_tag="t1735"
                     complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
     gene            REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
                     complement(844. .1797)
                     /gene="flgL"
     CDS             /locus_tag="t1736"
                     complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
     gene            NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
                     complement(1812. .3473)
                     /gene="flgK"
     CDS             /locus_tag="t1737"
                     complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
     gene            QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
                     complement(3538. .4488)
                     /gene="flgJ"
     CDS             /locus_tag="t1738"
                     complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
     gene            NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
                     complement(4488. .5591)
                     /gene="flgI"
     CDS             /locus_tag="t1739"
                     complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.7%;  Score 395.2;  DB 1;  Length 301983;
Best Local Similarity   91.7%;  Pred. No. 9.8e-103;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                 ||   | ||||  ||         |  || | ||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
     gene            KTAVVVQKQSHMAIRVTQR"
                     593. .1048
     CDS             /gene="agfA"
                     593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
     ORIGIN          IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"

  Query Match              85.4%;  Score 389.4;  DB 1;  Length 1048;
  Best Local Similarity    91.0%;  Pred. No. 2.7e-101;
  Matches  414;  Conservative     0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||   | ||||  ||       |  || | ||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACCGGCTAACCAGTATTA 455
              ||||||||||||||||  ||||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match              63.0%;  Score 287.2;  DB 6;  Length 361;
  Best Local Similarity    99.0%;  Pred. No. 8.4e-72;
  Matches  289;  Conservative     0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
```

```
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           61.8%;  Score 281.6;  DB 1;  Length 2889;
  Best Local Similarity  77.6%;  Pred. No. 4.3e-70;
  Matches  354;  Conservative    0;  Mismatches   99;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||| |||||||||||| |||||||||| |||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |      ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||   |  |||||| |||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||   ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||| ||||||||||  ||||| || || ||||||||  |||| |||| |
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              |      ||    | || | |||       | |  || | |||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571


RESULT 10
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMPTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN

  Query Match           50.3%;  Score 229.2;  DB 1;  Length 2920;
  Best Local Similarity  70.6%;  Pred. No. 5.9e-55;
  Matches  322;  Conservative    0;  Mismatches  128;  Indels    6;  Gaps    1;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| ||||||||||||      |||||| | || |||| || ||   ||| 
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || ||  ||| ||||   ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||     ||||| || ||   |||  ||| |||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||| ||| | ||||| |||  |||| |||| || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||| ||||| |||||||| || ||||| | ||||||| ||  ||| | | | |||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              |          | ||||  ||  |  |  | | | ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||| |||||||| || ||||||||   | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

RESULT 11
SEU53207

LOCUS	SEU53207 230 bp DNA linear BCT 01-MAY-1996
DEFINITION	Salmonella enteritidis SEF17 fimbrin (agfA) gene, partial cds.
ACCESSION	U53207
VERSION	U53207.1 GI:1293677
KEYWORDS	.
SOURCE	Salmonella enteritidis
ORGANISM	Salmonella enteritidis Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE	1 (bases 1 to 230)
AUTHORS	Cox,J.M., Eglezos,S. and Woolcock,J.B.
TITLE	Virulence of Salmonella enteritidis in chickens correlates with colony morphology and expression of SEF17 fimbriae
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 230)
AUTHORS	Cox,J.M., Eglezos,S. and Woolcock,J.B.
TITLE	Direct Submission
JOURNAL	Submitted (01-APR-1996) J.M. Cox, Food Science and Technology, The University of New South Wales, Sydney, NSW 2052, Australia

FEATURES Location/Qualifiers

source	1. .230
	/organism="Salmonella enteritidis"
	/mol_type="genomic DNA"
	/strain="SE30"
	/db_xref="taxon:592"
gene	/note="amplification product"
	1. .230
CDS	/gene="agfA"
	<1. .>230
	/gene="agfA"
	/codon_start=1
	/transl_table=11
	/product="SEF17 fimbrin"
	/protein_id="AAA98671.1"
	/db_xref="GI:1293678"
	/translation="GNHXGGGNSSGPDSTLSIYQYGSANAALALQSDARKSETTITQS GYGNGADVGQGADNSTIELTQNGFRNNATIDQ"

ORIGIN

Query Match 50.2%; Score 229; DB 1; Length 230;
Best Local Similarity 99.6%; Pred. No. 5.3e-55;
Matches 229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         88 GGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 147
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTAATCATAANGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 60

Qy        148 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACC 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACC 120

Qy        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 180

Qy        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 317
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTG 230
```

RESULT 12
ECCSGABDG

LOCUS	ECCSGABDG 4680 bp DNA linear BCT 07-JUL-2002
DEFINITION	E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION	X90754
VERSION	X90754.1 GI:1147558
KEYWORDS	csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene; orfC gene.
SOURCE	Escherichia coli
ORGANISM	Escherichia coli Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE	1
AUTHORS	Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
TITLE	Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12
JOURNAL	Mol. Microbiol. 18 (4), 661-670 (1995)
MEDLINE	96414468
PUBMED	8817489
REFERENCE	2 (bases 1 to 4680)
AUTHORS	Hammar,M.
TITLE	Direct Submission
JOURNAL	Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet, Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm, SWEDEN

FEATURES Location/Qualifiers

source	1. .4680
	/organism="Escherichia coli"
	/mol_type="genomic DNA"
	/strain="K12"
	/db_xref="taxon:562"
	/map="23.15 minutes"
source	1. .3140
	/organism="Escherichia coli"
	/mol_type="genomic DNA"
	/strain="MC4100"
	/db_xref="taxon:562"
source	3141. .4680
	/organism="Escherichia coli"
	/mol_type="genomic DNA"
	/strain="W3110"
	/db_xref="taxon:562"
repeat_region	<1. .>4680
	/insertion_seq="IS2 (partial)"
gene	complement(133. .966)
	/gene="csgG"
CDS	complement(133. .966)
	/gene="csgG"
	/codon_start=1

```
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match             49.8%;  Score 227.2;  DB 1;  Length 4680;
  Best Local Similarity   68.6%;  Pred. No. 2.3e-54;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || ||||||||||  ||||| || || || || ||    || |  | |
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||  |||  |||         |  |  |   ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 13
AE000205
```

```
LOCUS       AE000205                10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
     gene            /note="factor Sigma32; predicted +1 start at 1094983"
                     389. .1375
                     /gene="ycdU"
     CDS             /note="synonym: b1029"
                     389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
     protein_bind    /bound_moiety="DeoR predicted site"
                     1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
     promoter        /bound_moiety="DeoR predicted site"
                     1337. .1365
                     /gene="ycdU"
     gene            /note="factor Sigma70; predicted +1 start at 1096049"
                     1494. .1745
     CDS             /gene="b1030"
                     1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
```

```
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

  Query Match           49.8%;  Score 227.2;  DB 1;  Length 10346;
  Best Local Similarity   68.6%;  Pred. No. 2.5e-54;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| ||||||||  || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||

Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | | |
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                      ||  |||  |||       |  |  |  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448


RESULT 14
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
```

```
            Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
           Headed by:
            Name: Takashi Horiuchi
            Address: National Institute of Basic Biology, Okazaki, 444, Japan
            E-mail: kishori@nibb.ac.jp
           Information operator:
            Name: Hirotada Mori
            Address: NARA Institute of Science and Technology,
                     Ikoma, 630-01, Japan
            E-mail: hmori@gtc.aist-nara.ac.jp
           URL:
            The Japan E. coli genome database
            http:bsw3.aist-nara.ac.jp.
FEATURES    Location/Qualifiers
     source 1. .15047
            /organism="Escherichia coli K12"
            /mol_type="genomic DNA"
            /strain="K12"
            /db_xref="taxon:83333"
            /map="23.7-24.0 min"
            /clone="Kohara clone #231"
            /note="Nucleotide position 1097101-1112147 from the
            initiation site of ThrA (0 min.).~This clone is from
     CDS    Kohara lambda miniset library."
            <1. .276
            /note="unnamed protein product; ORF_ID:o230#5"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA35822.1"
            /db_xref="GI:4062605"
            /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
     CDS    WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
            1152. .2090
            /note="ORF_ID:o231#1
            similar to SwissProt Accession Number P45637"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical 33.0 kd protein in proB-proA
            intergenic region."
            /protein_id="BAA35823.1"
            /db_xref="GI:4062606"
            /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
            WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
            AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
            PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
            GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
     CDS    AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
            2145. .2882
            /note="unnamed protein product; ORF_ID:o231#2"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA35824.1"
            /db_xref="GI:4062607"
            /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
            EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
            PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
            SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
     CDS    ILNVSPRRLLNFLESRGMAPIAEFADL"
            2906. .3460
            /note="ORF_ID:o231#3
            similar to PIR Accession Number F64035"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein HI1543"
            /protein_id="BAA35825.1"
            /db_xref="GI:4062608"
            /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
            DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
            ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
     CDS    RTMAPLTRDAISAMWDELEEDSEE"
            3514. .4053
            /note="unnamed protein product; ORF_ID:o231#4"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA35826.1"
            /db_xref="GI:4062609"
            /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
            SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
            IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
     gene   LAARSAKTAHREQEIKNKA"
            complement(4117. .4950)
     CDS    /gene="csgG"
            complement(4117. .4950)
            /gene="csgG"
            /note="ORF_ID:o231#5
            similar to SwissProt Accession Number P52103"
            /codon_start=1
            /transl_table=11
            /product="Curli production assembly/transport component
            CsgG precursor."
            /protein_id="BAA35827.1"
            /db_xref="GI:4062610"
            /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
            PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
            NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
            IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
            VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
     gene   S"
            complement(4977. .5393)
     CDS    /gene="csgF"
            complement(4977. .5393)
            /gene="csgF"
            /note="ORF_ID:o231#6
            similar to SwissProt Accession Number P52104"
            /codon_start=1
            /transl_table=11
            /product="Curli production assembly/transport component
            CsgF precursor."
            /protein_id="BAA35828.1"
            /db_xref="GI:4062611"
            /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
            AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
     gene   DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
            complement(5418. .5807)
     CDS    /gene="csgE"
            complement(5418. .5807)
            /gene="csgE"
            /note="ORF_ID:o231#7
            similar to SwissProt Accession Number P52105"
            /codon_start=1
            /transl_table=11
            /product="Curli production assembly/transport component
            CsgE precursor."
            /protein_id="BAA35829.1"
            /db_xref="GI:4062612"
            /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
            RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
     gene   IQTEEALNRRQINQALLSTGDLAHDEF"
            complement(5812. .6462)
     CDS    /gene="csgD"
            complement(5812. .6462)
            /gene="csgD"
            /note="ORF_ID:o231#8
            similar to SwissProt Accession Number P52106"
            /codon_start=1
            /transl_table=11
            /product="Probable csgAB operon transcriptional regulatory
            protein."
            /protein_id="BAA35830.1"
            /db_xref="GI:4062613"
            /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
            RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
            HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
     gene   KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
            7217. .7672
```

```
                 /gene="csgB"
     CDS         7217. .7672

  Query Match           49.8%;  Score 227.2;  DB 1;  Length 15047;
  Best Local Similarity   68.6%;  Pred. No. 2.6e-54;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || |||||||||  ||||| || || || ||    || |  | |
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                 ||  ||| |||   |  |  |   ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||  |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 15
AX814811
LOCUS       AX814811                456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           49.5%;  Score 225.6;  DB 6;  Length 456;
  Best Local Similarity   68.4%;  Pred. No. 5.5e-54;
  Matches  312;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || |||||||||  ||||| || || || ||    || |  | |
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                 ||  ||| |||   |  |  |   ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||  |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


Search completed: March 18, 2004, 15:02:45
Job time : 2312.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64623	Aac64623 AgfA::PT3
	2	404.8	88.8	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	404.8	88.8	456	2	AAT74142	Aat74142 Salmonell
	4	404.8	88.8	456	3	AAC64617	Aac64617 Salmonell
	5	382.4	83.9	456	3	AAC64628	Aac64628 AgfA::PT3
	6	376	82.5	456	3	AAC64622	Aac64622 AgfA::PT3
	7	361.6	79.3	456	3	AAC64626	Aac64626 AgfA::PT3
	8	355.2	77.9	456	3	AAC64625	Aac64625 AgfA::PT3
	9	353.6	77.5	456	3	AAC64629	Aac64629 AgfA::PT3
	10	352	77.2	456	3	AAC64624	Aac64624 AgfA::PT3
	11	350.4	76.8	456	3	AAC64630	Aac64630 AgfA::PT3
	12	350.4	76.8	456	3	AAC64627	Aac64627 AgfA::PT3
	13	348.8	76.5	456	3	AAC64631	Aac64631 AgfA::PT3
	14	287.2	63.0	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	287.2	63.0	361	2	AAT74141	Aat74141 Salmonell
	16	225.6	49.5	456	3	AAC64619	Aac64619 Escherich
	17	225.6	49.5	456	9	ACF36153	Acf36153 E. coli C
	18	218	47.8	646	2	AAQ62647	Aaq62647 Fibronect
	19	156	34.2	369	2	AAQ62646	Aaq62646 FNB curli
	20	49	10.7	78	3	AAC64610	Aac64610 AgfA (SEF
c	21	48.4	10.6	78	3	AAC64609	Aac64609 AgfA (SEF
	22	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	23	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
c	24	44.8	9.8	2000	7	ADA71938	Ada71938 Rice gene
	25	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
c	26	41.6	9.1	78	3	AAC64605	Aac64605 SefA (SEF
	27	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
	28	40.6	8.9	78	3	AAC64606	Aac64606 SefA (SEF
	29	37.6	8.2	456	3	AAC64618	Aac64618 Salmonell
	30	37.4	8.2	456	3	AAC64620	Aac64620 Escherich
	31	35.4	7.8	100	7	ACD68807	Acd68807 E. coli K
c	32	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	33	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
	34	35	7.7	100	7	ACD68808	Acd68808 E. coli K
	35	34.8	7.6	1083	5	AAS76745	Aas76745 DNA encod
	36	34.4	7.5	2751	6	ABQ70750	Abq70750 Listeria
	37	34.2	7.5	2000	7	ADA71938	Ada71938 Rice gene
c	38	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	39	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	40	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	41	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	42	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	43	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c	44	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a
	45	33.4	7.3	78	9	ACF36151	Acf36151 E. coli C

ALIGNMENTS

RESULT 1
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.2e-138;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 2; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.2e-121;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||  | ||||  ||     |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 2;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.2e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||        |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

```
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.8%;  Score 404.8;  DB 3;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.2e-121;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| | |||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             83.9%;  Score 382.4;  DB 3;  Length 456;
  Best Local Similarity   89.9%;  Pred. No. 2.4e-114;
  Matches  410;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              |  |  |    ||  |  ||         |  || | ||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 6
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match            82.5%;  Score 376;  DB 3;  Length 456;
  Best Local Similarity  89.0%;  Pred. No. 2.9e-112;
  Matches  406;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||   ||||   |    | |  | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 79.3%; Score 361.6; DB 3; Length 456;
Best Local Similarity 87.1%; Pred. No. 1.5e-107;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||||||||||||| ||| |  |||||| || || || ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||   | ||||  ||          |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 8
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 77.9%; Score 355.2; DB 3; Length 456;
Best Local Similarity 86.2%; Pred. No. 1.8e-105;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||        |   ||  |   | | ||  | |    || | ||| ||||||
```

```
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 9
AAC64629
ID AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

```
  Query Match             77.5%;  Score 353.6;  DB 3;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 6.1e-105;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 2e-104;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              |||||| |  ||  ||   | | | |  | |   |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||   | |||| ||        |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 6.8e-104;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||  |     | | ||   |    |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||       |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36351.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 6.8e-104;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||  |  |      | ||  |    ||  |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||       ||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

 Query Match             76.5%;  Score 348.8;  DB 3;  Length 456;
 Best Local Similarity   85.3%;  Pred. No. 2.3e-103;
 Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
            ||||||  || |||  |   |  |||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||        |  || | ||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
            ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             63.0%;  Score 287.2;  DB 2;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 2.7e-83;
  Matches  289;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
               ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy         124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy         184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy         244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy         304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
               ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db         241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             63.0%;  Score 287.2;  DB 2;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 2.7e-83;
  Matches  289;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
               ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy         124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy         184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy         244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy         304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
               ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db         241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292


Search completed: March 18, 2004, 09:11:56
Job time : 289.405 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	404.8	88.8	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	287.2	63.0	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	33	7.2	7766	4	US-09-125-619-3	Sequence 3, Appli
c	6	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	7	31.2	6.8	1767	1	US-08-363-475-18	Sequence 18, Appl
c	8	31.2	6.8	2284	1	US-08-363-475-21	Sequence 21, Appl
	9	31	6.8	1750	1	US-08-415-823-1	Sequence 1, Appli
	10	31	6.8	1750	2	US-09-086-662-1	Sequence 1, Appli
c	11	30.6	6.7	1147	1	US-08-417-103-15	Sequence 15, Appl
c	12	30.6	6.7	1351	1	US-07-816-283-5	Sequence 5, Appli
c	13	30.6	6.7	1351	1	US-08-417-103-5	Sequence 5, Appli
c	14	30.6	6.7	1351	4	US-09-016-434-1303	Sequence 1303, Ap
c	15	30.6	6.7	1797	4	US-09-134-000C-1552	Sequence 1552, Ap
c	16	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	17	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	18	30.4	6.7	7183	4	US-09-081-149-9	Sequence 9, Appli
c	19	30.4	6.7	7183	4	US-09-081-149-10	Sequence 10, Appl
c	20	30.2	6.6	2007	4	US-09-643-597-153	Sequence 153, App
c	21	30.2	6.6	2007	4	US-09-480-884A-153	Sequence 153, App
c	22	30.2	6.6	2007	4	US-09-542-615A-153	Sequence 153, App
c	23	30.2	6.6	2007	4	US-09-606-421B-153	Sequence 153, App
c	24	30.2	6.6	2007	4	US-09-221-107-153	Sequence 153, App
c	25	30.2	6.6	2148	4	US-09-643-597-154	Sequence 154, App
c	26	30.2	6.6	2148	4	US-09-480-884A-154	Sequence 154, App
c	27	30.2	6.6	2148	4	US-09-542-615A-154	Sequence 154, App
c	28	30.2	6.6	2148	4	US-09-606-421B-154	Sequence 154, App
c	29	30.2	6.6	2148	4	US-09-221-107-154	Sequence 154, App
	30	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	31	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	32	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
c	33	29.8	6.5	3612	4	US-09-068-506-2	Sequence 2, Appli
c	34	29.8	6.5	3690	4	US-09-016-434-1234	Sequence 1234, Ap
c	35	29.8	6.5	4035	4	US-09-016-434-1369	Sequence 1369, Ap
c	36	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
c	37	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
c	38	29.4	6.4	759	2	US-08-853-659A-18	Sequence 18, Appl
c	39	29.4	6.4	918	2	US-08-853-659A-22	Sequence 22, Appl
c	40	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	41	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	42	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	43	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	44	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
c	45	29.4	6.4	9208	4	US-09-068-506-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.5e-131;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 63.0%; Score 287.2; DB 1; Length 361;
Best Local Similarity 99.0%; Pred. No. 1.8e-90;
Matches 289; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292
```

RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA AND ITS USE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,927
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,108
; FILING DATE: 06-AUG-1977
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1581
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1344 base pairs

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.13;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |  |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |    |||||| |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.13;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |  |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |    |||||| |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-09-125-619-3

  Query Match             7.2%;  Score 33;  DB 4;  Length 7766;
  Best Local Similarity   47.8%;  Pred. No. 0.77;
  Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||      ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |   |  |||||||||  | |||   |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |  || ||| |  |   |  ||| |         | |||  |||||  |  |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATATGAT 357
              || ||||       |||||
Db       4612 TTTGCTGTGAAGAGTAATGAT 4632


RESULT 6
5486473-3/c
;Patent No. 5486473
;    APPLICANT: FUKITA, HIROYUKI;YOSHIDA, IWAO;TAKAGI, MITSUO;
;MANABE, SADAO;FUKAI, KONOSUKE
;    TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
;    NUMBER OF SEQUENCES: 16
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,049
;      FILING DATE: 09-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 809,255
;      FILING DATE: 18-DEC-1991
;      APPLICATION NUMBER: 279,685
;      FILING DATE: 05-DEC-1988
;      APPLICATION NUMBER: 932,419
;      FILING DATE: 19-NOV-1986
;SEQ ID NO:3:
;      LENGTH: 1491
5486473-3

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 1491;
  Best Local Similarity   50.0%;  Pred. No. 1.3;
  Matches   78;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||  |  || ||| |  |  | |||    | | | || | | |  | | |  || |
Db        204 CAAGTCACTGACCGAAGCTAGGTAACAGTAACTGCGCACATCTGCGAGGTTGGCTGCTTC 145

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||    ||   ||     ||| ||    |    ||  |||||  || || |
Db        144 CATGTTCATCATTTTGACATCAATGGTTGGCTTGTCTTTTGACATTATGGTCACACAACT 85

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
              | |||  |    |||||   ||| |    | ||||
Db         84 ATCGCCTTCCAGTACCAGATCAACCCATGTAGCTCC 49
```

```
RESULT 7
US-08-363-475-18/c
; Sequence 18, Application US/08363475
; Patent No. 5516679
;  GENERAL INFORMATION:
;    APPLICANT:  Chiang, Shu-Jen
;    APPLICANT:  Burnett Jr., Willian V.
;    APPLICANT:  Tonzi, Sean M.
;    TITLE OF INVENTION:  PENICILLIN V AMIDOHYDROLASE GENE FROM
;    TITLE OF INVENTION:  FUSARIUM OXYSPORUM
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas R. Savitsky
;      STREET:  P.O. Box 4000
;      CITY:  Princeton
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/363,475
;      FILING DATE:  23-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savitsky, Thomas R.
;      REGISTRATION NUMBER:  31,661
;      REFERENCE/DOCKET NUMBER:  ON-0134
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 252 4956
;      TELEFAX:  (609) 252 4526
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1767 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1521
US-08-363-475-18

  Query Match             6.8%;  Score 31.2;  DB 1;  Length 1767;
  Best Local Similarity   55.6%;  Pred. No. 1.4;
  Matches   60;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        278 CTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATA 337
              | |  ||| ||||| |    || |||  | ||||||  |   ||  ||   |    | |
Db       1391 CCCCCAATCGTTTCTGGCGCAACGCCTTCCTCGACCCATTTCACAATAGCCAGAAGCACA 1332

Qy        338 TTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATG 385
              ||| | || |||    ||      |  ||| || | |||||| || |||
Db       1331 TTAGTATCTGCCGTCTTGCCGCCCATGTTTCCGGCGTTGTTTCCAATG 1284


RESULT 8
US-08-363-475-21/c
; Sequence 21, Application US/08363475
; Patent No. 5516679
;  GENERAL INFORMATION:
;    APPLICANT:  Chiang, Shu-Jen
;    APPLICANT:  Burnett Jr., Willian V.
;    APPLICANT:  Tonzi, Sean M.
;    TITLE OF INVENTION:  PENICILLIN V AMIDOHYDROLASE GENE FROM
;    TITLE OF INVENTION:  FUSARIUM OXYSPORUM
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas R. Savitsky
;      STREET:  P.O. Box 4000
;      CITY:  Princeton
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/363,475
;      FILING DATE:  23-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savitsky, Thomas R.
;      REGISTRATION NUMBER:  31,661
;      REFERENCE/DOCKET NUMBER:  ON-0134
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 252 4956
;      TELEFAX:  (609) 252 4526
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2284 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  241..1836
US-08-363-475-21

  Query Match             6.8%;  Score 31.2;  DB 1;  Length 2284;
  Best Local Similarity   55.6%;  Pred. No. 1.6;
  Matches   60;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        278 CTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATA 337
              | |  ||| ||||| |    || |||  | ||||||  |   ||  ||   |    | |
Db       1706 CCCCCAATCGTTTCTGGCGCAACGCCTTCCTCGACCCATTTCACAATAGCCAGAAGCACA 1647

Qy        338 TTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATG 385
              ||| | || |||    ||      |  ||| || | |||||| || |||
Db       1646 TTAGTATCTGCCGTCTTGCCGCCCATGTTTCCGGCGTTGTTTCCAATG 1599


RESULT 9
US-08-415-823-1
; Sequence 1, Application US/08415823
; Patent No. 5759538
;  GENERAL INFORMATION:
;    APPLICANT:  Donovan, William P.
;    APPLICANT:  Tan, Yuping
;    TITLE OF INVENTION:  BACILLUS THURINGIENSIS apr AND npr
;    TITLE OF INVENTION:  PROTEASE GENES AND ALKALINE PROTEASE DEFICIENT AND NEUTRAL
;    TITLE OF INVENTION:  PROTEASE DEFICIENT B.T. STRAINS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Panitch Schwarze Jacobs & Nadel, P.C.  c/o
;      ADDRESSEE:  A.S. Nadel
;      STREET:  1601 Market Street, 36th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103-2398
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,823
```

```
;       FILING DATE:
;       CLASSIFICATION:  435
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1750 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
US-08-415-823-1

  Query Match             6.8%;  Score 31;  DB 1;  Length 1750;
  Best Local Similarity   48.1%;  Pred. No. 1.6;
  Matches   88;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        254 CGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACC 313
              || | |||  |   |  | ||||   ||||    | |||  ||     | |||||
Db        449 CGTACAATTTTCAAAAGGTACTGTAAAAGAAAAAATAAAGAGCTATAAAAATAATCCAGA 508

Qy        314 AGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTG 373
               ||||||     | |||     ||||||  ||   |  | ||  | |     ||||
Db        509 TGTGGAATATGCAGAACCGAATTATTACGTTCACGCCTTTTGGACTCCAAACGACCCATA 568

Qy        374 TTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACA 433
              || | |      | | ||   |   |||  | | | | |   || ||||  | | ||||
Db        569 TTTTAATAATCAATACGGGTTACAAAAGATTCAAGCTCCACAAGCTTGGGATAGCCAACG 628

Qy        434 ACG 436
              | |
Db        629 AAG 631


RESULT 10
US-09-086-662-1
; Sequence 1, Application US/09086662
; Patent No. 5962264
;  GENERAL INFORMATION:
;    APPLICANT:  Donovan, William P.
;    APPLICANT:  Tan, Yuping
;    TITLE OF INVENTION:  BACILLUS THURINGIENSIS apr AND npr
;    TITLE OF INVENTION:  PROTEASE GENES AND ALKALINE PROTEASE DEFICIENT AND NEUTRAL
;    TITLE OF INVENTION:  PROTEASE DEFICIENT B.T. STRAINS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Panitch Schwarze Jacobs & Nadel, P.C.  c/o
;      ADDRESSEE:  A.S. Nadel
;      STREET:  1601 Market Street, 36th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103-2398
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/086,662
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/415,823
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1750 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
US-09-086-662-1

  Query Match             6.8%;  Score 31;  DB 2;  Length 1750;
  Best Local Similarity   48.1%;  Pred. No. 1.6;
  Matches   88;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        254 CGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACC 313
              || | |||  |   |  | ||||   ||||    | |||  ||     | |||||
Db        449 CGTACAATTTTCAAAAGGTACTGTAAAAGAAAAAATAAAGAGCTATAAAAATAATCCAGA 508

Qy        314 AGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTG 373
               ||||||     | |||     ||||||  ||   |  | ||  | |     ||||
Db        509 TGTGGAATATGCAGAACCGAATTATTACGTTCACGCCTTTTGGACTCCAAACGACCCATA 568

Qy        374 TTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACA 433
              || | |      | | ||   |   |||  | | | | |   || ||||  | | ||||
Db        569 TTTTAATAATCAATACGGGTTACAAAAGATTCAAGCTCCACAAGCTTGGGATAGCCAACG 628

Qy        434 ACG 436
              | |
Db        629 AAG 631


RESULT 11
US-08-417-103-15/c
; Sequence 15, Application US/08417103
; Patent No. 5723299
;  GENERAL INFORMATION:
;    APPLICANT:  Bell, Graeme I.
;    APPLICANT:  Yamada, Yuichiro
;    APPLICANT:  Seino, Susumu
;    TITLE OF INVENTION:  SOMATOSTATIN RECEPTORS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/417,103
;      FILING DATE:  05-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/816,283
;      FILING DATE:  01-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mark B.
;      REGISTRATION NUMBER:  37,259
;      REFERENCE/DOCKET NUMBER:  ARCD:144
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1147 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  15..1121
US-08-417-103-15

  Query Match             6.7%;  Score 30.6;  DB 1;  Length 1147;
  Best Local Similarity   48.1%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
              |||| |    |||   |  | |  |  |||||||  || ||||   | |  |  |||
```

```
Db        302 CAGACCCAGCATGAAGAGCTCATCTGCGATGGCCAGGTTGAGGATGTAAATGTTGGTGAT 243
Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
Db        242 GGTCTTCATCTTGGCATAGCGGAGGATGACATAAATGACAAGTGTGTTGCCACACAACCC 183
Qy        334 GATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393
Db        182 AATGATGCAGACCACAAAATAGATGAATGTGAGGACTGCATTGCTTGTCAGGTCATAGTA 123
Qy        394 C 394
Db        122 C 122
```

```
RESULT 12
US-07-816-283-5/c
; Sequence 5, Application US/07816283
; Patent No. 5436155
;  GENERAL INFORMATION:
;    APPLICANT:  Bell, Graeme I.
;    APPLICANT:  Yamada, Yuichiro
;    APPLICANT:  Seino, Susumu
;    TITLE OF INVENTION:  SOMATOSTATIN RECEPTORS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  PO Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/816,283
;      FILING DATE:  19911231
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McDaniel, C. Steven
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-787-1400
;      TELEFAX:  713-789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1351 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-07-816-283-5

  Query Match             6.7%;  Score 30.6;  DB 1;  Length 1351;
  Best Local Similarity   48.1%;  Pred. No. 2;
  Matches   87;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
Db        370 CAGACCCAGCATGAAGAGCTCATCTGCGATGGCCAGGTTGAGGATGTAAATGTTGGTGAT 311
Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
Db        310 GGTCTTCATCTTGGCATAGCGGAGGATGACATAAATGACAAGTGTGTTGCCACACAACCC 251
Qy        334 GATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393
Db        250 AATGATGCAGACCACAAAATAGATGAATGTGAGGACTGCATTGCTTGTCAGGTCATAGTA 191
Qy        394 C 394
Db        190 C 190
```

```
RESULT 13
US-08-417-103-5/c
; Sequence 5, Application US/08417103
; Patent No. 5723299
;  GENERAL INFORMATION:
;    APPLICANT:  Bell, Graeme I.
;    APPLICANT:  Yamada, Yuichiro
;    APPLICANT:  Seino, Susumu
;    TITLE OF INVENTION:  SOMATOSTATIN RECEPTORS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/417,103
;      FILING DATE:  05-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/816,283
;      FILING DATE:  01-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mark B.
;      REGISTRATION NUMBER:  37,259
;      REFERENCE/DOCKET NUMBER:  ARCD:144
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1351 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  83..1189
US-08-417-103-5

  Query Match             6.7%;  Score 30.6;  DB 1;  Length 1351;
  Best Local Similarity   48.1%;  Pred. No. 2;
  Matches   87;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
Db        370 CAGACCCAGCATGAAGAGCTCATCTGCGATGGCCAGGTTGAGGATGTAAATGTTGGTGAT 311
Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
Db        310 GGTCTTCATCTTGGCATAGCGGAGGATGACATAAATGACAAGTGTGTTGCCACACAACCC 251
Qy        334 GATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393
Db        250 AATGATGCAGACCACAAAATAGATGAATGTGAGGACTGCATTGCTTGTCAGGTCATAGTA 191
Qy        394 C 394
Db        190 C 190

RESULT 14
```

```
US-09-016-434-1303/c
; Sequence 1303, Application US/09016434
; Patent No. 6500938
;  GENERAL INFORMATION:
;    APPLICANT:  Janice Au-Young
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF SIGNALING
;    TITLE OF INVENTION:  PATHWAY GENE EXPRESSION
;    NUMBER OF SEQUENCES:  1490
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/016,434
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0002 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  1303:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1351 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GENBANK
;      CLONE:  g307435
US-09-016-434-1303

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 1351;
  Best Local Similarity   48.1%;  Pred. No. 2;
  Matches   87;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
              ||||| |    |||   |   | |   |  |||||||  || ||||   | | |  |||
Db        370 CAGACCCAGCATGAAGAGCTCATCTGCGATGGCCAGGTTGAGGATGTAAATGTTGGTGAT 311

Qy        274 CTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
                 |||  |||  |   |  ||| ||  |  |||  ||||    || | | |  ||
Db        310 GGTCTTCATCTTGGCATAGCGGAGGATGACATAAATGACAAGTGTGTTGCCACACAACCC 251

Qy        334 GATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393
               ||  | | | |  | ||   ||| |    |    | |  ||| |   || |  || | |
Db        250 AATGATGCAGACCACAAAATAGATGAATGTGAGGACTGCATTGCTTGTCAGGTCATAGTA 191

Qy        394 C 394
              |
Db        190 C 190


RESULT 15
US-09-134-000C-1552/c
; Sequence 1552, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1552
;   LENGTH: 1797
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-1552

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 1797;
  Best Local Similarity   65.2%;  Pred. No. 2.3;
  Matches   45;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||||||   | | |  ||   |||||||| ||||||  |||| |||  | | | | |
Db        721 CTGGTTTTTCTTCTGCTTCTTCATGAAATTGCACATATAAGCGCAATAATTCATTAATTG 662

Qy        421 GGTTTTGGC 429
              | | |||||
Db        661 GTTGTTGGC 653


Search completed: March 18, 2004, 20:20:31
Job time : 57.4971 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34.4	7.5	2751	15	US-10-398-221-3563	Sequence 3563, Ap
c 2	34.2	7.5	616	12	US-10-424-599-88945	Sequence 88945, A
c 3	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c 4	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c 5	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c 6	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c 7	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
c 8	34	7.5	1625	15	US-10-369-493-26980	Sequence 26980, A
c 9	33.2	7.3	1662	14	US-10-156-761-3483	Sequence 3483, Ap
10	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
11	33	7.2	7766	12	US-10-222-566-3	Sequence 3, Appli
12	33	7.2	7766	14	US-10-222-162-3	Sequence 3, Appli
13	33	7.2	7766	14	US-10-143-024-3	Sequence 3, Appli
14	33	7.2	368004	9	US-09-949-654-3	Sequence 3, Appli
15	32.8	7.2	1776	12	US-10-282-122A-24385	Sequence 24385, A
c 16	32.6	7.1	596	14	US-10-184-644-310	Sequence 310, App
c 17	32.6	7.1	596	14	US-10-184-634-310	Sequence 310, App
c 18	32.6	7.1	596	14	US-10-063-685-100	Sequence 100, App
19	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
c 20	32.2	7.1	708	12	US-10-282-122A-27953	Sequence 27953, A
21	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c 22	32	7.0	983	9	US-09-764-864-465	Sequence 465, App
23	31.8	7.0	750	14	US-10-184-644-104	Sequence 104, App
24	31.8	7.0	750	14	US-10-184-634-104	Sequence 104, App
c 25	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
26	31.8	7.0	4589	10	US-09-764-891-8845	Sequence 8845, Ap
27	31.8	7.0	4590	10	US-09-764-891-8846	Sequence 8846, Ap
28	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl
29	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
c 30	31.6	6.9	4830	12	US-10-282-122A-27869	Sequence 27869, A
c 31	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
32	31.4	6.9	269	9	US-09-923-876-5903	Sequence 5903, Ap
33	31.4	6.9	269	11	US-09-923-876-5903	Sequence 5903, Ap
c 34	31.4	6.9	1287	15	US-10-369-493-46950	Sequence 46950, A
35	31.4	6.9	2007	12	US-10-425-114-31950	Sequence 31950, A
36	31.4	6.9	2074	12	US-10-425-114-27886	Sequence 27886, A
c 37	31	6.8	487	14	US-10-060-036-4434	Sequence 4434, Ap
38	30.8	6.8	2886	9	US-09-801-368-131	Sequence 131, App
c 39	30.6	6.7	421	10	US-09-918-995-33997	Sequence 33997, A
c 40	30.6	6.7	1110	11	US-09-826-509-570	Sequence 570, App
c 41	30.6	6.7	1110	14	US-10-225-567A-313	Sequence 313, App
42	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
c 43	30.6	6.7	1351	15	US-10-305-720-1303	Sequence 1303, Ap
c 44	30.6	6.7	1510	15	US-10-292-798-677	Sequence 677, App
c 45	30.6	6.7	1608	9	US-09-815-242-3923	Sequence 3923, Ap

ALIGNMENTS

RESULT 1
US-10-398-221-3563
; Sequence 3563, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3563
; LENGTH: 2751
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3563

Query Match 7.5%; Score 34.4; DB 15; Length 2751;
Best Local Similarity 52.0%; Pred. No. 1.1;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy      238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
            |||||||   | |  ||| | |||| ||||||  ||      |       | || | | ||
Db      321 GATGTAGTGAAAGAAGCGAAAAATATTACTATCGACAAACCGTTTCCTCGTATGACATAT 380

Qy      298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGAT 357
            || |   | || || |   |     | | | ||  | ||||||  | | ||     | |
Db      381 AAAGAAGCAATGGATCGTTTTGGCAGTGATAAACCAGATATTCGTTTTGGTTTAGAATTA 440
```

```
Qy          358 CAGCTGGTTACCCGTGTTGTTACCCATG 385
Db              ||| || || ||| ||| ||| ||| |||
            441 CAAAATGTATCCGATGTCGTAAAAGATG 468

RESULT 2
US-10-424-599-88945/c
; Sequence 88945, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE:  2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88945
;  LENGTH: 616
;  TYPE: DNA
;  ORGANISM: Glycine max
;  FEATURE:
;  NAME/KEY: unsure
;  LOCATION: (1)..(616)
;  OTHER INFORMATION: unsure at all n locations
;  FEATURE:
;  OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

 Query Match
 Best Local Similarity   7.5%;  Score 34.2;  DB 12;  Length 616;
 Matches   81;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          101 GCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTA 160
Db              ||| || || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
            556 GGGCCCACAAGGGTAACGCCCCTGAATCCACGCCGGGTCCATGTGTGGATGATGATGTTG 497

Qy          161 ACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCG 220
Db              || ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
            496 ATGATGGTGATGATGAGGATGATGAGGCATGGGAGGTGGCACCGAAGGATAGTAACCGGG 437

Qy          221 GTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
Db              |||||||| ||| ||| ||| ||| ||| ||| |||
            436 GTAATGGTACTGAGGGTAATGCAGGGGAGGGTTCGGGTA 398

RESULT 3
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Qiu, Dewen
;  APPLICANT:  Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION:  RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE:  2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;  LENGTH: 1344
;  TYPE: DNA
;  ORGANISM: Erwinia amylovora
US-09-835-684-6

 Query Match
 Best Local Similarity   7.5%;  Score 34;  DB 9;  Length 1344;
 Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy           52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db              |||| || || || || || || || || || || || || || || || || || |||
            209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy          112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db              || || || | ||| || || || || || || || || || || || || || ||| |||
            149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy          172 GC 173
Db              ||
             89 GC 88

RESULT 4
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION:  PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE:  2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;  LENGTH: 1344
;  TYPE: DNA
;  ORGANISM: Erwinia amylovora
US-09-880-371-6

 Query Match
 Best Local Similarity   7.5%;  Score 34;  DB 9;  Length 1344;
 Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy           52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db              |||| || || || || || || || || || || || || || || || || || |||
            209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy          112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db              || || || | ||| || || || || || || || || || || || || || ||| |||
            149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy          172 GC 173
Db              ||
             89 GC 88

RESULT 5
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
;  APPLICANT: Fan, Hao
;  APPLICANT:  Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION:  THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE:  2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS  18
; SOFTWARE  PatentIn Ver  2 1
```

; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 6
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; APPLICANT: Oviedo, Agustin
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; TITLE OF INVENTION: FROM ORNAMENTAL PLANTS
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-010-390-6

Query Match 7.5%; Score 34; DB 14; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 7
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (EBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-10-441-736-5

Query Match 7.5%; Score 34; DB 15; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88
```

RESULT 8
US-10-369-493-26980/c
; Sequence 26980, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 26980
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-26980

Query Match 7.5%; Score 34; DB 15; Length 1625;
Best Local Similarity 61.1%; Pred. No. 1.1;
Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```
Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              || |||||  || ||   ||| |||  | || |    || ||  |  || || ||||||
Db        395 TCGGAAATTCTGTCAAGCTCGGCCACAGCAATGTCGCAAGCTTGGTCATAACCGTCGGCG 336

Qy        350 AATATGATCAGCTGGTTACCCGTGTTGTTA 379
              | | ||||  | ||| | ||| |||   ||
Db        335 ATTCTGATGGGGTGGATGCCCTTGTCAATA 306
```

RESULT 9
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

```
     ; APPLICANT: OMURA, SATOSHI
     ; APPLICANT:  IKEDA, HARUO
     ; APPLICANT:  ISHIKAWA, JUN
     ; APPLICANT:  HORIKAWA, HIROSHI
     ; APPLICANT:  SHIBA, TADAYOSHI
     ; APPLICANT:  SAKAKI, YOSHIYUKI
     ; APPLICANT:  HATTORI, MASAHIRA
     ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
     ; FILE REFERENCE: 249-262
     ; CURRENT APPLICATION NUMBER: US/10/156,761
     ; CURRENT FILING DATE:  2002-05-29
     ; PRIOR APPLICATION NUMBER: JP 2001-204089
     ; PRIOR FILING DATE: 2001-05-30
     ; PRIOR APPLICATION NUMBER: JP 2001-272697
     ; PRIOR FILING DATE: 2001-08-02
     ; NUMBER OF SEQ ID NOS: 15109
     ; SEQ ID NO 3483
     ;   LENGTH: 1662
     ;   TYPE: DNA
     ;   ORGANISM: Streptomyces avermitilis
     ;   FEATURE:
     ;   NAME/KEY: CDS
     ;   LOCATION: (1)..(1662)
     US-10-156-761-3483

       Query Match             7.3%;  Score 33.2;  DB 14;  Length 1662;
       Best Local Similarity   52.1%;  Pred. No. 2.1;
       Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

     Qy         50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
                   | || | ||  | |||||| | || |  | | || || || | ||| | |||||| |
     Db        590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531

     Qy        110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
                     ||| || | | |  ||  || | |  | || ||| | | |||
     Db        530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

     Qy        170 TTGCTCTGCAAAGCGATGCCCG 191
                   |||  || |  ||| | | |
     Db        470 TTGAACTTCTTGGCGGCGACGG 449


     RESULT 10
     US-10-156-761-1
     ; Sequence 1, Application US/10156761
     ; Publication No. US20030119018A1
     ; GENERAL INFORMATION:
     ;  APPLICANT: OMURA, SATOSHI
     ;  APPLICANT:  IKEDA, HARUO
     ;  APPLICANT:  ISHIKAWA, JUN
     ;  APPLICANT:  HORIKAWA, HIROSHI
     ;  APPLICANT:  SHIBA, TADAYOSHI
     ;  APPLICANT:  SAKAKI, YOSHIYUKI
     ;  APPLICANT:  HATTORI, MASAHIRA
     ;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
     ;  FILE REFERENCE: 249-262
     ;  CURRENT APPLICATION NUMBER: US/10/156,761
     ;  CURRENT FILING DATE:  2002-05-29
     ;  PRIOR APPLICATION NUMBER: JP 2001-204089
     ;  PRIOR FILING DATE: 2001-05-30
     ;  PRIOR APPLICATION NUMBER: JP 2001-272697
     ;  PRIOR FILING DATE: 2001-08-02
     ;  NUMBER OF SEQ ID NOS: 15109
     ,  SEQ ID NO 1
     ,    LENGTH: 9025608
     ,    TYPE: DNA
     ,    ORGANISM: Streptomyces avermitilis
     ,    FEATURE:
     ,    NAME/KEY: misc_feature
     ,    LOCATION: (4187715)
     ,    OTHER INFORMATION: a, t, c, g, other or unknown
     US-10-156-761-1

       Query Match             7.3%;  Score 33.2;  DB 14;  Length 9025608;
       Best Local Similarity   52.1%;  Pred. No. 1.8e+02;
       Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

     Qy         50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
                   | || | ||  | |||||| | || |  | | || || || | ||| | |||||| |
     Db   4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

     Qy        110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
                     ||| || | | |  ||  || | |  | || ||| | | |||
     Db   4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

     Qy        170 TTGCTCTGCAAAGCGATGCCCG 191
                   |||  || |  ||| | | |
     Db   4331901 TTGAACTTCTTGGCGGCGACGG 4331922


     RESULT 11
     US-10-222-566-3
     ; Sequence 3, Application US/10222566
     ; Publication No. US20040044192A1
     ; GENERAL INFORMATION:
     ;  APPLICANT: NORRIS, STEVEN J.
     ;  APPLICANT:  JING-REN, ZHANG
     ;  APPLICANT:  HARDHAM, JOHN M.
     ;  APPLICANT:  HOWELL, JERRILYN K.
     ;  APPLICANT:  BARBOUR, ALAN G.
     ;  APPLICANT:  WEINSTOCK, GEORGE M.
     ;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
     ;  FILE REFERENCE: UTSH:234USD3
     ;  CURRENT APPLICATION NUMBER: US/10/222,566
     ;  CURRENT FILING DATE:  2002-08-16
     ;  PRIOR APPLICATION NUMBER: 09/125,619
     ;  PRIOR FILING DATE: 1999-01-27
     ;  NUMBER OF SEQ ID NOS: 50
     ;  SOFTWARE: PatentIn Ver. 2.1
     ; SEQ ID NO 3
     ;   LENGTH: 7766
     ;   TYPE: DNA
     ;   ORGANISM: Borrelia burgdorferi
     ;   FEATURE:
     ;   NAME/KEY: modified_base
     ;   LOCATION: (127)
     ;   OTHER INFORMATION: R = A OR G
     US-10-222-566-3

       Query Match             7.2%;  Score 33;  DB 12;  Length 7766;
       Best Local Similarity   47.8%;  Pred. No. 5.8;
       Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

     Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
                   | || |||||| |||| |  |||| || ||| | | | | | | | |
     Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

     Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
                   | | | || || ||  | | |  | |||||||| | |||  ||| ||
     Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

     Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
                   |  || ||| |  |  |  ||| |  | ||| ||||| |  |  |
     Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

     Qy        337 ATTACTGTCGGCCAATATGAT 357
                   || ||||  |||||
     Db       4612 TTTGCTGTGAAGAGTAATGAT 4632


     RESULT 12
     US-10-222-162-3
     ; Sequence 3, Application US/10222162
     ; Publication No. US20030060618A1
     ; GENERAL INFORMATION:
     ;  APPLICANT: NORRIS, STEVEN J.
```

```
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD4
;  CURRENT APPLICATION NUMBER: US/10/222,162
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-162-3

  Query Match             7.2%;  Score 33;  DB 14;  Length 7766;
  Best Local Similarity   47.8%;  Pred. No. 5.8;
  Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |    ||||      ||  ||| |  | | |  | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |    |  | |||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |     |  ||| |           | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATATGAT 357
               || ||||       |||||
Db       4612 TTTGCTGTGAAGAGTAATGAT 4632


RESULT 13
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD1
;  CURRENT APPLICATION NUMBER: US/10/143,024
;  CURRENT FILING DATE:  2002-05-10
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-143-024-3

  Query Match             7.2%;  Score 33;  DB 14;  Length 7766;
  Best Local Similarity   47.8%;  Pred. No. 5.8;
  Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |    ||||      ||  ||| |  | | |  | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |    |  | |||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |     |  ||| |           | |||  |||||  |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATATGAT 357
               || ||||       |||||
Db       4612 TTTGCTGTGAAGAGTAATGAT 4632


RESULT 14
US-09-949-654-3
; Sequence 3, Application US/09949654
; Patent No. US20020127644A1
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000817
;  CURRENT APPLICATION NUMBER: US/09/949,654
;  CURRENT FILING DATE:  2001-09-12
;  PRIOR APPLICATION NUMBER: 60/231,572
;  PRIOR FILING DATE: 2000-09-11
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 368004
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(368004)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-654-3

  Query Match             7.2%;  Score 33;  DB 9;  Length 368004;
  Best Local Similarity   53.5%;  Pred. No. 48;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        328 AACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAA 387
              | ||| | | |||||| | || ||| |||||  | |  ||        | || |||| | |
Db     165120 ATCTCTGTTCTTACTGCCAGCAAATCTGATCTACAGAGTATAACATGTATTTCCCAAGCA 165179

Qy        388 ATGGCACATGCAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAAC 447
              | || |   ||  | || ||      |  |||  | || || |  |  ||  |||
Db     165180 CAGACAAAGCAAAAGGAAAAGGGCTAGGCAGGAGGTAGGGACACCAGCTGCTGGGGAAAC 165239

Qy        448 CAGTATTAA 456
               ||| || |
Db     165240 AAGTCTTCA 165248


RESULT 15
US-10-282-122A-24385
; Sequence 24385, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24385
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Listeria monocytogenes
US-10-282-122A-24385

Query Match 7.2%; Score 32.8; DB 12; Length 1776;
Best Local Similarity 51.4%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||| | | | |  ||  | ||||  || ||||||   |       | || | | | ||
Db        775 GATGTTGTCAAAGAAGCAAAAAATATCACCATTGAAAAACCGTTTCCTCGTATGACATAT 834

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGAT 357
              ||  |   | || || | |     | | | ||  | ||||||  | | ||   | |
Db        835 AAAGAAGCAATGGATCGTTTTGGCAGTGATAAACCGGATATTCGTTTTGGTTTAGAATTA 894

Qy        358 CAGCTGGTTACCCGTGTTGTTACCCATG 385
              ||    ||  ||  ||| || |   |||
Db        895 CAAAATGTATCCGATGTCGTAAAAGATG 922
```

Search completed: March 19, 2004, 03:49:29
Job time : 264.734 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-13	Sequence 13, Appl
2	404.8	88.8	456	6	US-08-233-642A-56	Sequence 56, Appl
3	404.8	88.8	456	23	US-09-543-407-1	Sequence 1, Appli
4	382.4	83.9	456	23	US-09-543-407-23	Sequence 23, Appl
5	376	82.5	456	23	US-09-543-407-11	Sequence 11, Appl
6	361.6	79.3	456	23	US-09-543-407-19	Sequence 19, Appl
7	355.2	77.9	456	23	US-09-543-407-17	Sequence 17, Appl
8	353.6	77.5	456	23	US-09-543-407-25	Sequence 25, Appl
9	352	77.2	456	23	US-09-543-407-15	Sequence 15, Appl
10	350.4	76.8	456	23	US-09-543-407-21	Sequence 21, Appl
11	350.4	76.8	456	23	US-09-543-407-27	Sequence 27, Appl
12	348.8	76.5	456	23	US-09-543-407-29	Sequence 29, Appl
13	287.2	63.0	361	6	US-08-233-642A-54	Sequence 54, Appl
14	227.2	49.8	477	17	US-09-252-691-172	Sequence 172, App
15	227.2	49.8	477	17	US-09-252-691C-172	Sequence 172, App
16	227.2	49.8	477	51	US-10-417-886-172	Sequence 172, App
17	225.6	49.5	456	23	US-09-543-407-3	Sequence 3, Appli
18	225.6	49.5	675	13	US-08-978-878-3	Sequence 3, Appli
19	225.6	49.5	675	31	US-09-741-873B-3	Sequence 3, Appli
20	223.2	48.9	453	92	US-60-352-946-1	Sequence 1, Appli
21	223.2	48.9	453	101	US-60-444-371-1	Sequence 1, Appli
22	181.6	39.8	396	13	US-08-978-878-1	Sequence 1, Appli
23	156.8	34.4	360	31	US-09-741-873B-1	Sequence 1, Appli
24	49	10.7	78	23	US-09-543-407-53	Sequence 1, Appli
c 25	48.4	10.6	78	23	US-09-543-407-52	Sequence 53, Appl
26	48	10.5	48	23	US-09-543-407-9	Sequence 52, Appl
c 27	44.8	9.8	2000	36	US-09-887-272A-5263	Sequence 9, Appli
c 28	41.6	9.1	78	23	US-09-543-407-48	Sequence 5263, Ap
29	40.8	8.9	100	45	US-10-146-492B-79	Sequence 48, Appl
30	40.6	8.9	78	23	US-09-543-407-49	Sequence 79, Appl
31	39.4	8.6	2732	51	US-10-437-963-37426	Sequence 49, Appl
32	37.6	8.2	456	23	US-09-543-407-2	Sequence 37426, A
33	37.6	8.2	522	17	US-09-252-691-170	Sequence 2, Appli
34	37.6	8.2	522	17	US-09-252-691C-170	Sequence 170, App
35	37.6	8.2	522	51	US-10-417-886-170	Sequence 170, App
36	37.4	8.2	456	23	US-09-543-407-4	Sequence 170, App
37	37.4	8.2	1304	51	US-10-437-963-19773	Sequence 4, Appli
38	35.4	7.8	100	45	US-10-146-492B-77	Sequence 19773, A
c 39	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 77, Appl
c 40	35.4	7.8	3412	1	PCT-US01-08631-24948	Sequence 24330, A
41	35	7.7	100	45	US-10-146-492B-78	Sequence 24948, A
42	34.8	7.6	1083	1	PCT-US01-08631-12549	Sequence 78, Appl
43	34.4	7.5	2751	50	US-10-398-221-3563	Sequence 12549, A
c 44	34.2	7.5	385	18	US-09-333-535A-288	Sequence 3563, Ap
c 45	34.2	7.5	385	25	US-09-615-606A-909	Sequence 288, App
						Sequence 909, App

ALIGNMENTS

```
RESULT 1
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 7.1e-135;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.8%;  Score 404.8;  DB 6;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.9e-118;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||     | ||||  ||          |  || | ||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.8%;  Score 404.8;  DB 23;  Length 456;
  Best Local Similarity   93.0%;  Pred. No. 1.9e-118;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||     | ||||  ||          |  || | ||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             83.9%;  Score 382.4;  DB 23;  Length 456;
  Best Local Similarity   89.9%;  Pred. No. 2.8e-111;
  Matches  410;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||  |||     ||  ||    || |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              |  |  |      ||  |  ||       |  ||  | ||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 82.5%; Score 376; DB 23; Length 456;
Best Local Similarity 89.0%; Pred. No. 3.2e-109;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
```

```
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||   ||||   |   | | |  ||  |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 1.3e-104;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||  |||    |  |||||  | |  |  ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 1.5e-102;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |    ||   |    | | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 4.8e-102;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||      |           ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 1.6e-101;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |    ||   ||    | | | |   | |     |   ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 5.1e-101;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||  |  |       | || |   || |   |    | ||
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             76.8%;  Score 350.4;  DB 23;  Length 456;
  Best Local Similarity   85.5%;  Pred. No. 5.1e-101;
  Matches  390;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                |  |||   |.      |  |  ||   |     |     |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

  Query Match            76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 1.7e-100;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              ||||||   || |||  |.    |                    |   | | ||    |
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                ||    | ||||  ||         |  || | ||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-642A-54

  Query Match            63.0%;  Score 287.2;  DB 6;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 8.7e-81;
  Matches  289;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATG 355
              |||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 292


RESULT 14
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match             49.8%;  Score 227.2;  DB 17;  Length 477;
  Best Local Similarity   68.6%;  Pred. No. 1.6e-61;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| |||||||||   | ||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||     |  |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  ||||||||| ||||||||   || |||| | |||||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||| |||||    ||| |||| | ||| |  ||    || |  |
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  |   |  ||   |||   |  | |     |||  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  ||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


RESULT 15
US-09-252-691C-172
; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-172

  Query Match             49.8%;  Score 227.2;  DB 17;  Length 477;
  Best Local Similarity   68.6%;  Pred. No. 1.6e-61;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| |||||||||   | ||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || |  || |||||||  |||||||||     |  |  | | || |||   | |  ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  ||||||||| ||||||||   || |||| | |||||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              |||||||||||||| |||||    ||| |||| | ||| |  ||    || |  |
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  |   |  ||   |||   |  | |     |||  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  ||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


Search completed: March 19, 2004, 03:01:49
Job time : 2650.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	225.6	49.5	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	181.6	39.8	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33	7.2	3990	5	US-09-830-230A-629	Sequence 629, App
	4	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
	5	31.4	6.9	1588	6	US-10-767-701-14522	Sequence 14522, A
	6	30.6	6.7	7096	6	US-10-021-698A-892	Sequence 892, App
	7	30.6	6.7	196083	6	US-10-021-698A-707	Sequence 707, App
c	8	30.2	6.6	2007	6	US-10-775-972-153	Sequence 153, App
c	9	30.2	6.6	2148	6	US-10-775-972-154	Sequence 154, App
	10	30	6.6	1355	6	US-10-045-674A-593	Sequence 593, App
	11	29.6	6.5	2159	6	US-10-767-701-13730	Sequence 13730, A
c	12	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
	13	29.6	6.5	318752	6	US-10-767-471-10553	Sequence 10553, A
	14	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
c	15	29.2	6.4	1790242	6	US-10-767-471-10805	Sequence 10805, A
	16	29	6.4	670	6	US-10-767-701-15726	Sequence 15726, A
c	17	29	6.4	861	6	US-10-767-701-3616	Sequence 3616, Ap
c	18	28.8	6.3	1857	1	PCT-US04-05654-2371	Sequence 2371, Ap
	19	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
	20	28.6	6.3	4634	6	US-10-791-799-10	Sequence 10, Appl
	21	28.4	6.2	1543	6	US-10-767-795-1782	Sequence 1782, Ap
	22	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	23	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	24	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
	25	28.2	6.2	798	6	US-10-100-683-2192	Sequence 2192, Ap
c	26	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
	27	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
	28	28	6.1	714	6	US-10-779-543-23469	Sequence 23469, A
c	29	27.8	6.1	390	5	US-09-796-692B-3647	Sequence 3647, Ap
	30	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
c	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	32	27.6	6.1	402	5	US-09-969-034-1464	Sequence 1464, Ap
c	33	27.6	6.1	564	6	US-10-767-701-16629	Sequence 16629, A
	34	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
	35	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	36	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	37	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
	38	27.6	6.1	10251	6	US-10-045-674A-582	Sequence 582, App
c	39	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
	40	27.4	6.0	819	6	US-10-417-884A-2244	Sequence 2244, Ap
	41	27.4	6.0	823	6	US-10-767-701-20007	Sequence 20007, A
c	42	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
	43	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
	44	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
	45	27.2	6.0	547	6	US-10-767-701-24545	Sequence 24545, A

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 49.5%; Score 225.6; DB 5; Length 675;
Best Local Similarity 68.4%; Pred. No. 1.5e-66;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873C
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873C-1

  Query Match             39.8%;  Score 181.6;  DB 5;  Length 396;
  Best Local Similarity   66.2%;  Pred. No. 8.7e-52;
  Matches  262;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396


RESULT 3
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Lyme Disease Vaccines
;  FILE REFERENCE: PB481US
;  CURRENT APPLICATION NUMBER: US/09/830,230A
;  CURRENT FILING DATE:  2001-09-27
;  PRIOR APPLICATION NUMBER: PCT/US98/12718
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/057,483
;  PRIOR FILING DATE: 1997-09-03
;  PRIOR APPLICATION NUMBER: 60/053,344
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/053,377
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/050,359
;  PRIOR FILING DATE: 1997-06-20
;  NUMBER OF SEQ ID NOS: 756
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
;   LENGTH: 3990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1139)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1143)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1148)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1210)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1244)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1247)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1250)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1251)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629
```

Query Match 7.2%; Score 33; DB 5; Length 3990;
Best Local Similarity 47.8%; Pred. No. 0.45;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  |||||| ||||    |    ||||     ||  ||| |  | | |  |  | |
Db       3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||     | |   |    | ||||||||||  | |||    |||  ||
Db       3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |     |  |||  |          | |||  |||||  |      |
Db       3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 3612

Qy        337 ATTACTGTCGGCCAATATGAT 357
              || ||||        |||||
Db       3613 TTTGCTGTGAAGAGTAATGAT 3633
```

RESULT 4
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
; APPLICANT: Neubauer, Antonie
; APPLICANT: Ziegler, Christina
; TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
; FILE REFERENCE: 1/1372
; CURRENT APPLICATION NUMBER: US/10/624,149A
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 60/403,282
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: DE 10233064
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: DE 10317008
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
; LENGTH: 150223
; TYPE: DNA
; ORGANISM: Equine herpesvirus 1
; PUBLICATION INFORMATION:
; AUTHORS: Telford,E.A.
; AUTHORS: Watson, M.S.
; AUTHORS: McBride, K.
; AUTHORS: Davison, A.J.
; TITLE: The DNA sequence of equine herpesvirus-1
; JOURNAL: Virology
; VOLUME: 189
; ISSUE: 1
; PAGES: 304-316
; DATE: JUL-1992
; DATABASE ACCESSION NUMBER: NC 001491, NCBI
; DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 3.8;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              || ||  ||  |  | || ||  | ||| ||| |   |  ||| |||   | |||   |||
Db      95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy        233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
              |       |  |||| ||  ||||||     | || |  | | | ||| |    |  |||
Db      95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy        293 GAAATAATGCCACCATCGACCAGTGG 318
              || |   |     || ||| | || |
Db      95662 GAGAGCTTATTCACAACGAGCTGTTG 95687
```

RESULT 5
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
; LENGTH: 1588
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

Query Match 6.9%; Score 31.4; DB 6; Length 1588;
Best Local Similarity 44.5%; Pred. No. 1;
Matches 125; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

```
Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
              | |  |||| |        | ||| | |  |      | ||||| |||  |  |  |  |
Db        455 GGTGGTGGCGGCTTCGGTGGTGGTGGCTATGGCGGTGGCGGCGGTGGCTATGGTGGTGGC 514

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCT 159
              |||||  ||| || ||   |||   |      | |   |||          ||||  |  |
Db        515 GGCGGTGGCTATGGTGGTGGCTATGGTGGCAACTACGGCAACAGGGGTGGCGGTGGCTAT 574

Qy        160 AACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
              |  ||   ||| |     |  | | ||| ||  | |||  |   | | ||  |     |
Db        575 GGTGGTGGTGTTGGTGATTATGGAGCTGCAGGTGGAGCTGGAGGCAACTTTGCTGCTGGA 634

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACT 279
              |||  |   | |   |||          |  ||||| || | | ||  |  || |    ||
Db        635 GGTAGTAACAGCTTTGCCAGCAGTAACTTCGGTGCTGACAGTGGTTTTGGTGGAAACCCT 694

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAA 320
              | | ||  |   |  | ||||   || | ||  | |
Db        695 GCTGGTAGCTTTGGCGGTGGTTCCACTGGCGCCGAGAAGGA 735
```

RESULT 6
US-10-021-698A-892
; Sequence 892, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4044US1
; CURRENT APPLICATION NUMBER: US/10/021,698A
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/211,749
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 6160
; SOFTWARE: Patentin 2.1
; SEQ ID NO 892
; LENGTH: 7096
; TYPE: DNA

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (7085)..(7087)
;   OTHER INFORMATION: a, t, c or g
US-10-021-698A-892

  Query Match              6.7%;  Score 30.6;  DB 6;  Length 7096;
  Best Local Similarity   48.6%;  Pred. No. 3.8;
  Matches   84;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        278 CTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATA 337
              ||| | ||      |||| ||   |  |  |  |||     || | |||| | | ||
Db       2964 CTCTCCAAGGGCCAGGAAACAACATCTTAAGATAAAAGTCACTGGCCATAAACAACCTTA 3023

Qy        338 TTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 397
              ||  |    ||| |  || || ||  ||| ||| ||| |    ||| |       |
Db       3024 TTCAAGGAAAATAATCTCGTCTGCAGGGCACCTGTGGTGTGAAGGCTGAGACCTGTGAGT 3083

Qy        398 CAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |  |  | || || ||   ||     ||| || ||||| ||| |
Db       3084 TCAGCAACAGAATCGGCCAAGTGGGGGCTGCAGTAAACCCCTCGGCTTACCGG 3136


RESULT 7
US-10-021-698A-707
; Sequence 707, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 707
;   LENGTH: 196083
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (256)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (89496)..(89595)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (141214)..(141313)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (182468)..(182567)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (193506)..(193605)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (194799)..(194898)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (195354)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (195460)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (195644)
;   OTHER INFORMATION: a, t, c or g
US-10-021-698A-707

  Query Match              6.7%;  Score 30.6;  DB 6;  Length 196083;
  Best Local Similarity   48.6%;  Pred. No. 17;
  Matches   84;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        278 CTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATA 337
              ||| | ||      |||| ||   |  |  |  |||     || | |||| | | ||
Db      45914 CTCTCCAAGGGCCAGGAAACAACATCTTAAGATAAAAGTCACTGGCCATAAACAACCTTA 45973

Qy        338 TTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 397
              ||  |    ||| |  || || ||  ||| ||| ||| |    ||| |       |
Db      45974 TTCAAGGAAAATAATCTCGTCTGCAGGGCACCTGTGGTGTGAAGGCTGAGACCTGTGAGT 46033

Qy        398 CAAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |  |  | || || ||   ||     ||| || ||||| ||| |
Db      46034 TCAGCAACAGAATCGGCCAAGTGGGGGCTGCAGTAAACCCCTCGGCTTACCGG 46086


RESULT 8
US-10-775-972-153/c
; Sequence 153, Application US/10775972
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C21
;  CURRENT APPLICATION NUMBER: US/10/775,972
;  CURRENT FILING DATE:  2004-02-10
;  NUMBER OF SEQ ID NOS: 563
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-972-153

  Query Match              6.6%;  Score 30.2;  DB 6;  Length 2007;
  Best Local Similarity   45.8%;  Pred. No. 2.9;
  Matches  104;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        140 TTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTG 199
              ||   ||| |   | |   || | ||  | | | |   |  |  |  |||| |  |
Db       1416 TTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTAAGTTAAAAGTACTTAAAAGTT 1357

Qy        200 AAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              | |  |||   ||  || ||| |  ||| |   |  ||||  |  |  | |||
Db       1356 CACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCTGCCTAGGGGAATTTCAGCATT 1297

Qy        260 ATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGA 319
              | |  |  ||| || ||| |   | |  ||||||    ||  | |  ||||||
Db       1296 AGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATAATCTGCCCTCCCATTAGTGGT 1237

Qy        320 ACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTT 366
              | | | || |||| | |   | | |  || | | ||    |||
Db       1236 AGGAGAGAAGCTCCAAGTTCCGAGGCTTCTCATTTTACCAAACAGTT 1190


RESULT 9
```

US-10-775-972-154/c
; Sequence 154, Application US/10775972
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C21
; CURRENT APPLICATION NUMBER: US/10/775,972
; CURRENT FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 563
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 2148
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-972-154

Query Match 6.6%; Score 30.2; DB 6; Length 2148;
Best Local Similarity 45.8%; Pred. No. 3;
Matches 104; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy        140 TTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTG 199
              ||   ||| |   |  |   || | ||  | |  | |   |    | |  | ||||  |
Db       1549 TTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTAAGTTAAAAGTACTTAAAAGTT 1490

Qy        200 AAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              |  |  |||   ||  || ||| |  ||| |    |    |||| |   |  | |||
Db       1489 CACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCTGCCTAGGGGAATTTCAGCATT 1430

Qy        260 ATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGA 319
              |  |     |   ||| || |||  |    | |    |||||    ||  | |  |||||
Db       1429 AGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATAATCTGCCCTCCCATTAGTGGT 1370

Qy        320 ACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTT 366
              | |  | || |||| |  |   | |  |  || | ||   |||
Db       1369 AGGAGAGAAGCTCCAAGTTCCGAGGCTTCTCATTTTACCAAACAGTT 1323
```

RESULT 10
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)

US-10-045-674A-593

Query Match 6.6%; Score 30; DB 6; Length 1355;
Best Local Similarity 25.1%; Pred. No. 2.8;
Matches 84; Conservative 80; Mismatches 170; Indels 0; Gaps 0;

```
Qy         57 GGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTC 116
              :| :||:| :  :     : |  ::||:||:||:: : |   :|| ||: :: |  |:
Db        768 KGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGG 827

Qy        117 CGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCT 176
              :||: : |      :   |:: :  : | | :| :| :|  ||    ||   :||:
Db        828 YGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAA 887

Qy        177 GCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGC 236
              |: |:| :     : |:  :|        | : | | :    :| |:| :|| ||:
Db        888 YAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTGAYGCYAARGGYAA 947

Qy        237 CGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAA 296
              :|   ::   | : | ||: |: || ||  :         :   | :||:  :   :
Db        948 RYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYATYGGYGAYGTYWS 1007

Qy        297 TAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              :  :    | |  :| : |  |   :  |:  :  |:| :| :  :  |:     :||:   |
Db       1008 YGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKAATTCYCARATGGC 1067

Qy        357 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 390
              :||: | | :  : |:| :  ::: | :  :|||
Db       1068 YCARGTYGGWGAYGGKGAYAAYWSWCCKYTWATG 1101
```

RESULT 11
US-10-767-701-13730
; Sequence 13730, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 13730
; LENGTH: 2159
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS49140_1
US-10-767-701-13730

Query Match 6.5%; Score 29.6; DB 6; Length 2159;
Best Local Similarity 48.8%; Pred. No. 4.8;
Matches 80; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy        288 TTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGG 347
              || ||  | | ||| ||||   |||  |   |  || |||||    ||||  | |   |
Db        682 TTACACTATTCATGACACCGAGGACACGGAAACTGCCAAAAAAGATGATACAAGTTATGA 741

Qy        348 CCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAAT 407
              |||  | |          |       ||| | |  |  ||   |  ||| |||   ||| ||
Db        742 CCAGCAGGGGAGCAAGCCCCTGTGTGATTTCTCAAATTTTAGAGCAAATGTGTGCGAGAT 801

Qy        408 GGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGT 451
              |    || | ||| |  ||  |   || || || |  | |  ||
Db        802 GCGTGGTGATGTTAGAGTTCACCCAAATGCTACAGAAATCATGT 845
```

RESULT 12
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:

```
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
;   LENGTH: 24900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(24900)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

  Query Match                 6.5%;  Score 29.6;  DB 6;  Length 24900;
  Best Local Similarity   52.4%;  Pred. No. 15;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||  | || ||  | ||| || |  | | || || |  | | |||||  |
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |  |||||  |  | | |||||  | ||||| ||| | | | |  |||  |||
Db        9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929

Qy        327 AAAC 330
              ||| |
Db        9928 AATC 9925


RESULT 13
US-10-767-471-10553
; Sequence 10553, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10553
;   LENGTH: 318752
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(318752)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10553

  Query Match                 6.5%;  Score 29.6;  DB 6;  Length 318752;
  Best Local Similarity   52.4%;  Pred. No. 46;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |  |  | |  || ||||||  || ||||||| ||  || || ||  | | | | ||
Db     254839 TGTATACCACACCCAGAAACCTTGCAGAAATCATCATGACAGAGATGAGATAATCTCAAA 254898

Qy        327 AAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGA 386
              |  || ||| |||       | || ||  ||||||    || ||| | |  | | |
Db     254899 ACTTTCTGATTTTATGTGGCGATCATTTGCCATGCTGGTAGAAGGTTTTATGAGAACTTA 254958

Qy        387 AATG 390
              ||||
Db     254959 AATG 254962


RESULT 14
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match                 6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 2.3;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||     |  |||| | |  |  | |  || || |||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
              |||||||||||||| ||     |||||||||  |  ||  | || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 15
US-10-767-471-10805/c
; Sequence 10805, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10805
;   LENGTH: 1790242
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1790242)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10805

  Query Match                 6.4%;  Score 29.2;  DB 6;  Length 1790242;
  Best Local Similarity   53.5%;  Pred. No. 1e+02;
  Matches   61;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        266 CTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              || ||   |||  |||||||   ||||  | | ||||  | |  ||   ||| |
Db     369207 CTGTTCTGTTGATGCAGAATGAAATCAGGTAAACTGCCCTAAGCTTCCTCAAGATCTACT 369148

Qy        326 AAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTA 379
              || |||  | |  | ||||  |||  ||||         |  |||| | | || | |
Db     369147 CAAGCTCAGTTTATCCTGTTAGCCTATATCTCTTTTTCTTTACACATTTTTTAA 369094


Search completed: March 19, 2004, 03:10:11
Job time : 58.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-13
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	226.2	49.6	680	12	BJ618688	BJ618688 BJ618688
	2	36	7.9	523	12	BJ334624	BJ334624 BJ334624
c	3	35.6	7.8	744	14	CF446143	CF446143 EST682488
c	4	34.8	7.6	549	28	BZ295150	BZ295150 CG1320.r1
	5	34.8	7.6	594	12	BJ335653	BJ335653 BJ335653
	6	34.8	7.6	813	28	BZ204853	BZ204853 CH230-390
	7	34.6	7.6	540	14	CB210843	CB210843 OML01123
	8	34.6	7.6	556	9	AV976624	AV976624 AV976624
	9	34.4	7.5	539	13	BQ141266	BQ141266 NF017G03P
	10	34.4	7.5	560	10	AW686303	AW686303 NF036D06N
	11	34.4	7.5	627	10	AW573882	AW573882 EST316473
	12	34.2	7.5	456	12	BJ331339	BJ331339 BJ331339
	13	34.2	7.5	524	12	BJ367372	BJ367372 BJ367372
	14	34.2	7.5	547	12	BJ333958	BJ333958 BJ333958
	15	34.2	7.5	553	12	BJ336903	BJ336903 BJ336903
	16	34.2	7.5	569	12	BJ334047	BJ334047 BJ334047
	17	34.2	7.5	574	12	BJ330328	BJ330328 BJ330328
	18	34.2	7.5	607	12	BJ365766	BJ365766 BJ365766
	19	34.2	7.5	625	12	BJ396922	BJ396922 BJ396922
	20	34.2	7.5	629	12	BJ335378	BJ335378 BJ335378
	21	34.2	7.5	635	12	BJ397391	BJ397391 BJ397391
	22	34.2	7.5	645	12	BJ392753	BJ392753 BJ392753
	23	34.2	7.5	670	12	BJ336637	BJ336637 BJ336637
	24	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	25	34	7.5	234	10	BB581987	BB581987 BB581987
c	26	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	27	34	7.5	512	10	BB757100	BB757100 BB757100
	28	33.8	7.4	298	14	F19524	F19524 HSPD04128 H
c	29	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	30	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	31	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
	32	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	33	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
c	34	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	35	33.6	7.4	811	28	BZ469000	BZ469000 BOOAO58TF
	36	33.4	7.3	402	13	BU767674	BU767674 SJEBBF08
c	37	33.4	7.3	505	6	AL809250	Al809250 Triticum
	38	33.4	7.3	536	13	BU776718	BU776718 SJEDCA03
	39	33.4	7.3	546	13	BU780674	BU780674 SJEFZH02
c	40	33.4	7.3	699	12	BI959328	BI959328 HVSMEn001
	41	33.4	7.3	715	13	BU444683	BU444683 603765715
	42	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
	43	33.2	7.3	629	9	AU295325	AU295325 AU295325
	44	33.2	7.3	788	9	AU139321	AU139321 AU139321
c	45	33.2	7.3	1646	10	BF579120	BF579120 602096054

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 49.6%; Score 226.2; DB 12; Length 680;
Best Local Similarity 68.4%; Pred. No. 3.5e-53;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
              || ||  | || ||||||||||  ||||| || || || || ||    || |  | |
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
                  ||  |||  |||         |  |  |   ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
BJ334624
LOCUS BJ334624 523 bp mRNA linear EST 05-MAR-2002
DEFINITION BJ334624 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda47o06 5', mRNA sequence.
ACCESSION BJ334624
VERSION BJ334624.1 GI:19164754
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 523)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES Location/Qualifiers
source 1. .523
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47o06"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Query Match 7.9%; Score 36; DB 12; Length 523;
Best Local Similarity 55.5%; Pred. No. 30;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||  ||   | |    ||| |||||  ||||| | | | |   ||||    ||  |||
Db        398 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGANGACAATAAT 457

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     |||||| |   |   |  ||||
Db        458 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAATAANGATGATGATGA 516
```

RESULT 3
CF446143/c
LOCUS CF446143 744 bp mRNA linear EST 04-SEP-2003
DEFINITION EST682488 normalized cDNA library of onion Allium cepa cDNA clone ACAHU74, mRNA sequence.
ACCESSION CF446143
VERSION CF446143.1 GI:34468845
KEYWORDS EST.
SOURCE Allium cepa (onion)
ORGANISM Allium cepa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae; Allium.
REFERENCE 1 (bases 1 to 744)
AUTHORS Havey,M.J., Cheung,F., Van Aken,S., Utterback,T. and Town,C.D.
TITLE Expressed Sequence Tags from a normalized library of mixed onion tissues (Allium cepa)
JOURNAL Unpublished (2003)
COMMENT Contact: Havey MJ
Department of Horticulture
USDA-ARS and University of Wisconsin
1575 Linden Drive, Madison, WI 53706, USA
Tel: 608-262-1830
Fax: 608-262-4743
Email: mjhavey@facstaff.wisc.edu
TIGR sequence name ACAHU74TR. For more information:
http://haveylab.hort.wisc.edu
Seq primer: CAG GAA ACA GCT ATG ACC.
FEATURES Location/Qualifiers
source 1. .744
/organism="Allium cepa"
/mol_type="mRNA"
/cultivar="Red Creole(bulbs), unknown(callus), Ebano & Texas Legend(roots)"
/db_xref="taxon:4679"
/clone="ACAHU74"
/tissue_type="Callus, roots, and young bulbs"
/clone_lib="normalized cDNA library of onion"
/note="Vector: pCMVSport6.1-ccdb (Invitrogen); Site_1: EcoRV (5'); Site_2: NotI (3'); Equal molar amounts of mRNA from callus, roots, and young bulbs were combined to synthesize the library. Normalization to enrich for low-copy transcripts was performed by proprietary techniques of Invitrogen."
ORIGIN

Query Match 7.8%; Score 35.6; DB 14; Length 744;
Best Local Similarity 47.7%; Pred. No. 46;

```
  Matches  104;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy        234 CGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAG 293
              ||  | || | || |        |||  | || |  | | ||||||   |  | | || |
Db        382 CGAAGCTGAAAGCAACAAACTTTATAGCAATATTGATCATCTGACTGCTATGGATATCCG 323

Qy        294 AAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATA 353
              ||||||    |  | |  | | |||| | ||||||  |  | |  |||| ||    |  |
Db        322 AAATAACAATAGTAACAGCAAATGGATCTCTAAAAGGTGTGTTGATACTATCAATAATGA 263

Qy        354 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCG 413
               || | |   |  | |     | ||    |||||   |  |||      |  | || | |
Db        262 AGAGCGGAGAGAAAGCAACTCTTTTTATGATGAATCTGTTCATAATGTTGGCAAGGAGGG 203

Qy        414 TCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGT 451
                ||       |||||||    ||||     |||
Db        202 AAGATTTAACAACTTCAACAACAAATCGGCCTCGAAGT 165


RESULT 4
BZ295150/c
LOCUS       BZ295150                549 bp    DNA     linear   GSS 31-OCT-2002
DEFINITION  CG1320.r1 Candida glabrata Random Genomic Library Candida glabrata
            genomic clone CG1320, genomic survey sequence.
ACCESSION   BZ295150
VERSION     BZ295150.1  GI:24436228
KEYWORDS    GSS.
SOURCE      Candida glabrata
  ORGANISM  Candida glabrata
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
  TITLE     Evidence from comparative genomics for a complete sexual cycle in
            the 'asexual' pathogenic yeast Candida glabrata
  JOURNAL   Genome Biol. 4 (2), R10 (2003)
  MEDLINE   22508158
   PUBMED   12620120
COMMENT     Contact: Wong S
            Department of Genetics, Smurfit Institute
            Trinity College Dublin
            Dublin 2, Ireland
            Tel: 353 1 6082319
            Fax: 353 1 6798558
            Email: swong@tcd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Candida glabrata"
                     /mol_type="genomic DNA"
                     /strain="CBS 138"
                     /db_xref="taxon:5478"
                     /clone="CG1320"
                     /clone_lib="Candida glabrata Random Genomic Library"
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 28;  Length 549;
  Best Local Similarity   52.0%;  Pred. No. 68;
  Matches   78;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        222 TTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCA 281
              ||||  ||     ||  ||  |   ||||||    |     | | |  | .   ||  |    ||
Db        291 TTATAGTTTTTGCTCCTACAAAGGCCAAATTTTCAGTGAAAAAAGCTGAAAATTTTTGCA 232

Qy        282 GAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTAC 341
              |||    |  |||       |  || ||   ||   | || ||| |  | |||| |||| |
Db        231 GAAAATTGACAGTGTGTGTATCAGCAAACACATCTAGATCGCAACCAGCTCCTATATCAG 172

Qy        342 TGTCGGCCAATATGATCAGCTGGTTACCCG 371
              |||| |||||  |  |||   | | || |||
Db        171 TGTCTGCCAAGAATATCCTTTTGATAACCG 142


RESULT 5
BJ335653
LOCUS       BJ335653                594 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ335653 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda51h06 5', mRNA sequence.
ACCESSION   BJ335653
VERSION     BJ335653.1  GI:19165783
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda51h06"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 12;  Length 594;
  Best Local Similarity   55.5%;  Pred. No. 70;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||  ||   | |     ||| |||||   |||||  | | | |   ||||    ||  |||
Db        223 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 282

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || ||  | || |    ||   ||||||     ||||||  |   |   |  ||||
Db        283 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAATAAAGATGATGATGA 341


RESULT 6
BZ204853
LOCUS       BZ204853                813 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-390D21, genomic survey sequence.
ACCESSION   BZ204853
VERSION     BZ204853.1  GI:23862905
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-390D21.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
```

```
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 390  row: D  column: 21
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-390D21"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             7.6%;  Score 34.8;  DB 28;  Length 813;
  Best Local Similarity   50.0%;  Pred. No. 80;
  Matches   87;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              ||| |   | | | | || |   |    ||   || | ||  |  |   ||  |  ||
Db        548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              || |   || ||       || |  | |  |||||  |     |||||  ||  |   |
Db        608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAA 351
              ||   ||  |||||||| ||| |    | |||| | |   | |||| ||| ||
Db        668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAA 721


RESULT 7
CB210843
LOCUS       CB210843                 540 bp    mRNA    linear   EST 05-FEB-2003
DEFINITION  OML01123 Oryza minuta HybriZAP-2.1 XR library Oryza minuta cDNA 5',
            mRNA sequence.
ACCESSION   CB210843
VERSION     CB210843.1  GI:28256934
KEYWORDS    EST.
SOURCE      Oryza minuta
  ORGANISM  Oryza minuta
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Shin,J.S.
  TITLE     Oryza minuta HybriZAP-2.1 XR library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Jeong Sheop Shin
            Plant Molecular Genetics
            Graduate School of Biotechnology, University of Korea
            136-701 Anam-dong 5/1 Seoul, Korea
            Tel: 00 82 2 3290 3430
            Fax: 00 82 2 927 9028
            Email: jsshin@kuccnx.korea.ac.kr.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Oryza minuta"
                     /mol_type="mRNA"
                     /db_xref="taxon:63629"
                     /dev_stage="4-weeks after germination"
                     /clone_lib="Oryza minuta HybriZAP-2.1 XR library"
                     /note="Organ: immature leaf"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 14;  Length 540;
  Best Local Similarity   50.3%;  Pred. No. 77;
  Matches   85;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        225 TGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAA 284
              |||   | | |   || |||    || |  ||||||  ||    |||||||  |
Db        140 TGGAGCCTGTGGTCCTGGAGGTTTTGGAGATGATAATGATAAATCTGAACTGTTTATTGT 199

Qy        285 TGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGT 344
              | |||||  ||  ||||| ||   |||    | ||| |||   ||  |  ||| |  | |
Db        200 TCGTTTCCAAAGAAATGCAACATACGAAGCATTGAAAGCTGCGAAAACAAATAATCTTAT 259

Qy        345 CGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 393
              |      |    | |  |||   || ||   ||||   | |||||
Db        260 GGAGGCTCTCCAAAGACAGCATACTACTGATGCCTCCCAGCTAGTGGCA 308


RESULT 8
AV976624
LOCUS       AV976624                 556 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  AV976624 Nori Satoh unpublished cDNA library, egg Ciona
            intestinalis cDNA clone cieg42j24 5', mRNA sequence.
ACCESSION   AV976624
VERSION     AV976624.1  GI:19466390
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cieg42j24"
                     /tissue_type="whole animal"
                     /dev_stage="egg"
                     /clone_lib="Nori Satoh unpublished cDNA library, egg"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 9;  Length 556;
  Best Local Similarity   47.1%;  Pred. No. 78;
  Matches  106;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy        222 TTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCA 281
              | |||  || |     |||||   | | ||  |  |  | || |  ||     ||| |
Db         67 TGATGTCAAAGCTCATGATGTCACACCTGATGATGTCACCACTAATGATGTCACCACTAA 126

Qy        282 GAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTAC 341
              |||    ||  ||| ||| |||||   |  |    || ||| |  |   |  || ||
Db        127 TGATGTCACCACTAATGATGTCACCACTAATGATGTCAAAGCTGATGATGTCACTACTAA 186

Qy        342 TGTCGGCCAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAG 401
              ||  | |  | |||| |  |   |||   || ||| || | |||     |    | | |
Db        187 TGATGTCACACCTGATAATGTCACTACTAATGATGTCACACCTGATGATGTCACTACTAA 246

Qy        402 CGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAA 446
              |  |      | | | ||    |   ||  | | |||  ||||
```

```
Db        247 TGATGTCAAAGCTGATGATGTCACTACTAATGATGTCACACCTAA 291


RESULT 9
BQ141266
LOCUS       BQ141266                539 bp    mRNA    linear   EST 26-APR-2002
DEFINITION  NF017G03PH1F1024 Phoma-infected Medicago truncatula cDNA clone
            NF017G03PH 5', mRNA sequence.
ACCESSION   BQ141266
VERSION     BQ141266.1  GI:20277392
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
            Gonzales,R.A., Bell,C.J., Inman,J.T., Waugh,M.E., Sullivan,J.P.,
            May,G.D. and Paiva,N.L.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula Phoma-infected library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Paiva NL
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7317
            Fax: 580 221 7380
            Email: nlpaiva@noble.org
            Insert Length: 539   Std Error: 0.00
            Plate: 017  row: G  column: 03
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF017G03PH"
                     /tissue_type="leaf"
                     /dev_stage="Pathogen-induced, young trifoliate"
                     /clone_lib="Phoma-infected"
                     /note="Vector: pBluescript SK(-); Young trifoliate leaves
                     of Medicago truncatula were excised and dip-inoculated in
                     a spore suspension of Phoma medicaginis, and incubated in
                     humid dishes. Pools of leaves were harvested at 0, 15, and
                     30 minutes and 1, 2, 3, 6, 14, 24, 48, 72, and 96, hours,
                     and used to prepare total RNA. cDNA was prepared from
                     polyA+ enriched, pooled samples of equivalent amounts of
                     total RNA from each sample. The cDNA was directionally
                     ligated into the Uni-Zap XR vector (Stratagene) and
                     packaged using the Gigapack III Gold packaging extracts.
                     Phagemids containing cDNA inserts were in vivo excised
                     from the recombinant Uni-ZAP XR vector using ExAssist
                     helper phage and the E. coli strain XL1-Blue MRF'
                     (Stratagene). Excised plasmids were plated using SOLR
                     cells."
ORIGIN

  Query Match             7.5%;  Score 34.4;  DB 13;  Length 539;
  Best Local Similarity   46.0%;  Pred. No. 88;
  Matches  116;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy        129 AACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||  || |  |  | | | | |||||  |  ||| ||    | |        ||  ||| ||
Db        233 AATTTTAAAAACATCTGACTCCGGTTACAATAATGCATACCCTTTGAATGAAGACGAGGC 292

Qy        189 CCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCA 248
              |||  |  |      ||    | | ||       |       | || | ||   |   |
Db        293 ACGTGATCCACCTGAGATTCCTCCACAATTGCAAAACACCTTGGTCGGCTATCCTGCTAA 352

Qy        249 GGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCAT 308
```

```
              |  |    |  |||| || ||  | || ||||||| |   |   |||| ||      ||
Db        353 TGTCGATAGTTTTAGTTCTCTTCCATTGCCTCAGAACGCAATTCTAAATCATCTTTATAT 412

Qy        309 CGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTAC 368
              ||  |  |  |  | |||  || |    |  || | ||  |   | |||   |  |||
Db        413 TGAGAATAGAGAGCCACAAAGATCTGTGGTGGCTTTGGGATACACTCATCGCTTTCGTAC 472

Qy        369 CCGTGTTGTTAC 380
              |    |||||||
Db        473 CAAATTTGTTAC 484


RESULT 10
AW686303
LOCUS       AW686303                560 bp    mRNA    linear   EST 20-DEC-2000
DEFINITION  NF036D06NR1F1000 Nodulated root Medicago truncatula cDNA clone
            NF036D06NR 5', mRNA sequence.
ACCESSION   AW686303
VERSION     AW686303.2  GI:11931130
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
            Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W.,
            May,G.D. and Paiva,N.L.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula nodulated root library
  JOURNAL   Unpublished (2000)
COMMENT     On Apr 14, 2000 this sequence version replaced gi:7561039.
            Contact: Paiva NL
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7317
            Fax: 580 221 7380
            Email: nlpaiva@noble.org
            Insert Length: 665   Std Error: 0.00
            Plate: 036  row: D  column: 06
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF036D06NR"
                     /tissue_type="root"
                     /dev_stage="Pooled developmental"
                     /clone_lib="Nodulated root"
                     /note="Vector: Lambda Zap; Four-week old Rhizobium
                     meliloti-inoculated Medicago truncatula roots, containing
                     a mixture of young and old roots and nodules."
ORIGIN

  Query Match             7.5%;  Score 34.4;  DB 10;  Length 560;
  Best Local Similarity   46.0%;  Pred. No. 89;
  Matches  116;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy        129 AACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||  || |  |  | | | | |||||  |  ||| ||    | |        ||  ||| ||
Db        230 AATTTTAAAAACATCTGACTCCGGTTACAATAATGCATACCCTTTGAATGAAGACGAGGC 289

Qy        189 CCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCA 248
              |||  |  |      ||    | | ||       |       | || | ||   |   |
Db        290 ACGTGATCCACCTGAGATTCCTCCACAATTGCAAAACACCTTGGTCGGCTATCCTGCTAA 349

Qy        249 GGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCAT 308
              |  |    |  |||| || ||  | || ||||||| |   |   |||| ||      ||
Db        350 TGTCGATAGTTTTAGTTCTCTTCCATTGCCTCAGAACGCAATTCTAAATCATCTTTATAT 409
```

```
Qy        309 CGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTAC 368
              ||  |  |  |  |  |||  || |  |  || | ||  |  | |||  |  |||
Db        410 TGAGAATAGAGAGCCACAAAGATCTGTGGTGGCTTTGGGATACACTCATCGCTTTCGTAC 469

Qy        369 CCGTGTTGTTAC 380
              |    |||||||
Db        470 CAAATTTGTTAC 481
```

RESULT 11
AW573882
LOCUS AW573882 627 bp mRNA linear EST 07-SEP-2000
DEFINITION EST316473 GVN Medicago truncatula cDNA clone pGVN-49K8, mRNA sequence.
ACCESSION AW573882
VERSION AW573882.1 GI:7238615
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 627)
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name:M252780e
TIGR sequence name:MTCAJ64TK
More information is available at. .
'http://chrysie.tamu.edu/medicago'
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES Location/Qualifiers
source 1. .627
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-49K8"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XLOLR"
/clone_lib="GVN"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells."
ORIGIN

Query Match 7.5%; Score 34.4; DB 10; Length 627;
Best Local Similarity 46.0%; Pred. No. 93;
Matches 116; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy        129 AACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||  || |  |  | | | | ||||| |  ||| ||    | |       ||  ||| ||
Db        177 AATTTTAAAAACATCTGACTCCGGTTACAATAATGCATACCCTTTGAATGAAGACGAGGC 236

Qy        189 CCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCA 248
              ||| |  |       ||    | | ||          |       | || | ||   |   |
Db        237 ACGTGATCCACCTGAGATTCCTCCACAATTGCAAAACACCTTGGTCGGCTATCCTGCTAA 296

Qy        249 GGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCAT 308
              | |     | |||| || ||  | || ||||||| |   |   |||| ||       ||
Db        297 TGTCGATAGTTTTAGTTCTCTTCCATTGCCTCAGAACGCAATTCTAAATCATCTTTATAT 356

Qy        309 CGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTAC 368
              ||  |  |  |  |  |||  || |  |  || | ||  |  | |||  |  |||
Db        357 TGAGAATAGAGAGCCACAAAGATCTGTGGTGGCTTTGGGATACACTCATCGCTTTCGTAC 416

Qy        369 CCGTGTTGTTAC 380
              |    |||||||
Db        417 CAAATTTGTTAC 428
```

RESULT 12
BJ331339
LOCUS BJ331339 456 bp mRNA linear EST 05-MAR-2002
DEFINITION BJ331339 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda34m22 5', mRNA sequence.
ACCESSION BJ331339
VERSION BJ331339.1 GI:19161469
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 456)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .456
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda34m22"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 456;
Best Local Similarity 55.5%; Pred. No. 93;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||  ||   | |    ||| |||||   |||||  | | | |   ||||    ||  |||
Db        264 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 323

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||    ||||||  |   |   |  ||||
Db        324 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGATGA 382
```

RESULT 13
BJ367372
LOCUS BJ367372 524 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ367372 Dictyostelium discoideum cDNA library, CF Dictyostelium discoideum cDNA clone ddc42p08 5', mRNA sequence.
ACCESSION BJ367372
VERSION BJ367372.1 GI:19276674

```
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc42p08"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 524;
  Best Local Similarity   55.5%;  Pred. No. 99;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        181 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 240

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     |||||| |   |   |  ||||
Db        241 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGATGA 299


RESULT 14
BJ333958
LOCUS       BJ333958                547 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ333958 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda44j16 5', mRNA sequence.
ACCESSION   BJ333958
VERSION     BJ333958.1  GI:19164088
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda44j16"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 547;
  Best Local Similarity   55.5%;  Pred. No. 1e+02;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        188 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 247

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     |||||| |   |   |  ||||
Db        248 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGATGA 306


RESULT 15
BJ336903
LOCUS       BJ336903                553 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ336903 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda55g13 5', mRNA sequence.
ACCESSION   BJ336903
VERSION     BJ336903.1  GI:19167033
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda55g13"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 553;
  Best Local Similarity   55.5%;  Pred. No. 1e+02;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        184 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 243

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGA 356
              |||  || || | || |    ||   ||||||     |||||| |   |   |  ||||
Db        244 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGATGA 302


Search completed: March 18, 2004, 20:11:27
Job time : 2035.55 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	403.2	88.4	456	6	I44909	I44909 Sequence 58
2	403.2	88.4	2067	1	SEU43280	U43280 Salmonella
3	395.2	86.7	5103	1	STAJ2301	AJ002301 Salmonell
4	395.2	86.7	22411	1	AE008749	AE008749 Salmonell
5	393.6	86.3	254050	1	AL627269	AL627269 Salmonell
c 6	393.6	86.3	301983	1	AE016840	AE016840 Salmonell
7	387.8	85.0	1048	1	STAGFBA	AJ000514 Salmonell
8	284.2	62.3	361	6	I44908	I44908 Sequence 56
9	278.4	61.1	2889	1	CSP515700	AJ515700 Citrobact
10	229.2	50.3	2920	1	CFR515701	AJ515701 Citrobact
11	225.4	49.4	1711	1	AF275733	AF275733 Escherich
12	225.4	49.4	10190	1	AE005315	AE005315 Escherich
13	225.4	49.4	327773	1	AP002554	AP002554 Escherich
14	224	49.1	4680	1	ECCSGABDG	X90754 E.coli csgG
15	224	49.1	10346	1	AE000205	AE000205 Escherich
16	224	49.1	15047	1	D90741	D90741 Escherichia
17	223.8	49.1	306358	1	AE016759	AE016759 Escherich
18	222.4	48.8	456	6	AX814811	AX814811 Sequence
19	222.4	48.8	648	1	ECOCSGAA	L04979 Escherichia
20	208	45.6	2883	1	ESA515702	AJ515702 Enterobac
21	203.2	44.6	230	1	SEU53207	U53207 Salmonella
22	186.4	40.9	10370	1	AE015131	AE015131 Shigella
23	186.4	40.9	292504	1	AE016981	AE016981 Shigella
c 24	163.4	35.8	437	1	AF237726	AF237726 Shigella
25	93.2	20.4	19201	1	D90742	D90742 Escherichia
26	50	11.0	1212	1	ECO131756	AJ131756 Escherich
27	48.2	10.6	78	6	AX814809	AX814809 Sequence
28	41.6	9.1	72	6	AX814798	AX814798 Sequence
c 29	36.2	7.9	1425	1	AF343445	AF343445 Lactobaci
c 30	36.2	7.9	302156	3	AC116977	AC116977 Dictyoste
31	35.6	7.8	13430	1	AE000409	AE000409 Escherich
32	35.6	7.8	108366	2	AC020660	AC020660 Homo sapi
33	35.6	7.8	110000	1	ECOUW67_2	Continuation (3 of
c 34	35.6	7.8	110000	2	AC145807_0	AC145807 Silurana
c 35	35.6	7.8	196539	2	AC145912	AC145912 Gallus ga
c 36	35.4	7.8	168267	2	AC146009	AC146009 Pan trogl
37	35.2	7.7	163619	2	AC119493	AC119493 Rattus no
38	34.6	7.6	11160	1	AE008209	AE008209 Agrobacte
c 39	34.6	7.6	11476	1	AE009405	AE009405 Agrobacte
40	34.6	7.6	125020	9	AF429315	AF429315 Homo sapi
41	34.2	7.5	1458	1	AF173043	AF173043 Orientia
c 42	34.2	7.5	179906	2	AC119683	AC119683 Rattus no
43	34.2	7.5	268098	2	AC133387	AC133387 Rattus no
44	34.2	7.5	299708	2	AC110963	AC110963 Rattus no
45	34	7.5	956	8	AK058943	AK058943 Oryza sat

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             88.4%;  Score 403.2;  DB 6;  Length 456;
  Best Local Similarity   92.8%;  Pred. No. 2.7e-111;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match             88.4%;  Score 403.2;  DB 1;  Length 2067;
  Best Local Similarity   92.8%;  Pred. No. 3.1e-111;
```

```
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |   | |     |   ||   ||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match           86.7%;  Score 395.2;  DB 1;  Length 5103;
  Best Local Similarity  91.7%;  Pred. No. 9.1e-109;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |  |   | |     |   ||   ||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
     RBS                complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene               complement(2580. .4088)
                        /gene="STM1128"
     CDS                complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS                complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene               complement(4413. .5099)
                        /gene="STM1129"
     CDS                complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
     RBS                complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene               5599. .6759
                        /gene="STM1130"
     CDS                5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAPSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene               6793. .7497
                        /gene="STM1131"
     RBS                6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           86.7%;  Score 395.2;  DB 1;  Length 22411;
  Best Local Similarity  91.7%;  Pred. No. 1e-108;
  Matches  418;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAININSSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAARERKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             86.3%;  Score 393.6;  DB 1;  Length 254050;
  Best Local Similarity   91.4%;  Pred. No. 3.9e-108;
  Matches  417;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |  |   | |     |   ||   |||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             86.3%;  Score 393.6;  DB 1;  Length 301983;
Best Local Similarity   91.4%;  Pred. No. 4e-108;
  Matches  417;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||     | | | |   |   | |      |   ||   |||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             85.0%;  Score 387.8;  DB 1;  Length 1048;
  Best Local Similarity   90.8%;  Pred. No. 1.4e-106;
  Matches  413;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |  | |     |   ||  ||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||||  |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             62.3%;  Score 284.2;  DB 6;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 4.3e-75;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
```

```
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           61.1%;  Score 278.4;  DB 1;  Length 2889;
  Best Local Similarity   77.2%;  Pred. No. 3e-73;
  Matches  352;  Conservative    0;  Mismatches  101;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||||| |||||||||| |||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||    | || || ||       | | |  | |  |   |     |          ||||||
Db       2179 GG---TTCTGTTCCGCAATGGGGCGGCGGCGGTGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||   |  |||||| |||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||    ||||  || | ||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || ||||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || ||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571


RESULT 10
CFR515701
LOCUS       CFR515701               2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN

  Query Match           50.3%;  Score 229.2;  DB 1;  Length 2920;
  Best Local Similarity   70.6%;  Pred. No. 2.8e-58;
  Matches  322;  Conservative    0;  Mismatches  128;  Indels    6;  Gaps    1;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||   |  ||  ||      | |    |  ||||     |    |  |  |||
Db       2183 GGTGTTGTTCCGCAATGGG------GCGGCAATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||   ||| |||| ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||    ||||| || ||   |||  ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||||| ||| | ||||| |||   ||||  ||||  || |||  || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||  ||||| |||||||| || |||||  | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||| ||| |||||||||||||| || ||||| |||| ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || |||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

RESULT 11
AF275733
LOCUS AF275733 1711 bp DNA linear BCT 15-MAY-2001
DEFINITION Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and CsgA protein (csgA) genes, complete cds.
ACCESSION AF275733
VERSION AF275733.1 GI:14039399
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1711)
AUTHORS Uhlich,G.A., Keen,J.E. and Elder,R.O.
TITLE Mutations in the csgD promoter associated with variations in curli expression in certain strains of Escherichia coli O157:H7
JOURNAL Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
MEDLINE 21218556
PUBMED 11319125
REFERENCE 2 (bases 1 to 1711)
AUTHORS Uhlich,G.A., Keen,J.E. and Elder,R.O.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES Location/Qualifiers
source 1. .1711
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="43895 Red Variant"
/serotype="O157:H7"
/db_xref="taxon:562"
gene 757. .1212
/gene="csgB"
CDS 757. .1212
/gene="csgB"
/codon_start=1
/transl_table=11
/product="CsgB protein"
/protein_id="AAK53211.1"
/db_xref="GI:14039400"
/translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
gene 1253. .1711
/gene="csgA"
CDS 1253. .1711
/gene="csgA"
/codon_start=1
/transl_table=11
/product="CsgA protein"
/protein_id="AAK53212.1"
/db_xref="GI:14039401"
/translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
ORIGN

```
Query Match             49.4%;  Score 225.4;  DB 1;  Length 1711;
Best Local Similarity   69.7%;  Pred. No. 3.8e-57;
Matches  320;  Conservative    0;  Mismatches  136;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | ||||||||||||| | || || || ||||||||
Db       1253 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 1312

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCA---TGAAATGGCACATGCATCC 117
              || ||   |  ||||    |    | |||  | ||  |||    ||      |   ||     |
Db       1313 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 1372

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              |||||| | |||  | |||  ||||||||||||||    | ||||  |||| ||||||||
Db       1373 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 1432

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||   |||||  |||||  |||||     ||  |||||||||    |||  |||||||| ||
Db       1433 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 1492

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||||  || |||||  | |||  | ||  | ||  || |||||  ||   |||  ||  | ||
Db       1493 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 1552

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              |  |||||  | ||  |||||||||| |||| ||||  ||||  || ||     |||| |||
Db       1553 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 1612

Qy        358 GGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAG 417
              ||    ||||  ||  |||| | ||||| |||||| |||||| ||| |  ||    ||||
Db       1613 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 1672

Qy        418 GTTGGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||| ||||| |||||||| || ||| | ||||| |||
Db       1673 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 1711
```

RESULT 12
AE005315
LOCUS AE005315 10190 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139 of 155.
ACCESSION AE005315 AE005174
VERSION AE005315.1 GI:12514572
KEYWORDS .
SOURCE Escherichia coli O157:H7 EDL933
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10190)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,

```
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10190
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            617. .1072
                     /gene="csgB"
                     /note="synonym: Z1675"
     CDS             617. .1072
                     /gene="csgB"
                     /function="structural component; Cell exterior
                     constituents: Surface structures"
                     /note="Residues 1 to 151 of 151 are 100.00 pct identical
                     to residues 1 to 151 of 151 from Escherichia coli K-12
                     Strain MG1655: B1041"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAG55787.1"
                     /db_xref="GI:12514573"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1113. .1571
                     /gene="csgA"
                     /note="synonym: Z1676"
     CDS             1113. .1571
                     /gene="csgA"
                     /function="regulator; Cell exterior constituents: Surface
                     structures"
                     /note="Residues 1 to 152 of 152 are 96.71 pct identical to
                     residues 1 to 151 of 151 from Escherichia coli K-12 Strain
                     MG1655: B1042"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin major subunit, coiled surface structures;
                     cryptic"
                     /protein_id="AAG55788.1"
                     /db_xref="GI:12514574"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
     gene            1630. .1962
                     /gene="csgC"
                     /note="synonym: Z1677"
     CDS             1630. .1962
                     /gene="csgC"
                     /function="putative structure; Not classified"
                     /note="Residues 1 to 110 of 110 are 98.18 pct identical to
                     residues 1 to 110 of 110 from Escherichia coli K-12 Strain
                     MG1655: B1043"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="AAG55789.1"
                     /db_xref="GI:12514575"
                     /translation="MNALLLLAALSSQITFNTTQQGDMYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     gene            2083. .2394
                     /gene="Z1678"
     CDS             2083. .2394
                     /gene="Z1678"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 103 of 103 are 98.05 pct identical to
                     residues 1 to 103 of 103 from Escherichia coli K-12 Strain
                     MG1655: B1044"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55790.1"
                     /db_xref="GI:12514576"
                     /translation="MFRPFLNSLMLGSLFFPFIAIAGSTAQGGVIHFYGQIVEPACDV
                     STQSSPVEMNCPQNGSVPGKTYSSKALMSGNVKNAQIASVKVQYLDKQKKLAVMNIEY
                     N"
     gene            2489. .3022
                     /gene="Z1679"
     CDS             2489. .3022
                     /gene="Z1679"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 177 of 177 are 100.00 pct identical
                     to residues 1 to 177 of 177 from Escherichia coli K-12
                     Strain MG1655: B1045"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyprotein"
                     /protein_id="AAG55791.1"
                     /db_xref="GI:12514577"
                     /translation="MKTRIHVVQGDITKLAVDVIVNAANPSLMGGGGVDGAIHRAAGP
                     ALLDACLKVRQQQGDCPTGHAVITLAGDLPAKAVVHTVGPVWRGGEQNEDQLLQDAYL
                     NSLRLVAANSYTSVAFPAISTGVYGYPRAAAAEIAVKTVSEFITRHALPEQVYFVCYD
                     EENAHLYERLLTQQGDE"
     gene            2964. .4445
                     /gene="ymdC"
                     /note="synonym: Z1680"
     CDS             2964. .4445
                     /gene="ymdC"
                     /function="putative enzyme; Not classified"
                     /note="Residues 1 to 493 of 493 are 99.59 pct identical to
                     residues 1 to 493 of 493 from Escherichia coli K-12 Strain
                     MG1655: B1046"
                     /codon_start=1
                     /transl_table=11
                     /product="putative synthase"
                     /protein_id="AAG55792.1"
                     /db_xref="GI:12514578"
                     /translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPGQCGLFPL
                     EKSLDAFAARYRLAEMSEHTLDVQYYIWQDDMSGRLLFSALLAAAKRGVRVRLLLDDN
                     NTPGLDDILRLLDSHPRIEVRLFNPFSFRLLRPLGYITDFSRLNRRMHNKSFTVDGVV
                     TLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWYCKSVSPLQQVLDV
                     PEGEMADRIELPASWHNDAMTHRYLRKMESSPPINHLVDGTLPLIWAKTRLLSDDPAK
                     GEGKAKRHSLLPQRLFDIMGSPSERIDIISSYFVPTRAGVAQLLRMVRKXVKIAILTN
                     SLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREQSSTLHDRGITGNSGASLHAKT
                     FSIDGKTVFIGSFNFDPRSTLLNTEMGFVIESETLAQLIDKRFIQSQYDAAWQLRLDR
                     WGRINWVDRHAKKEIVLKKEPATSFWKRVMVRLASILPVEWLL"
     gene            complement(4453. .5610)
                     /gene="ymdD"
                     /note="synonym: Z1681"
     CDS             complement(4453. .5610)
                     /gene="ymdD"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 385 of 385 are 99.22 pct identical to
                     residues 1 to 385 of 385 from Escherichia coli K-12 Strain
                     MG1655: B1047"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55793.1"
```

```
                     /db_xref="GI:12514579"
                     /translation="MNPVPAQREYFXDSIRAWLMLLGIPFHISLIYSSHTWHVNSAEP
                     SLWLTLFNDFIHSFRMQVFFVISGYFSYMLFLRYPLKKWWKVRVERVGIPMLTAIPLL
                     TLPQFIMLQYVKGKAESWPGLSLYDKYNTLAWELISHLWFLLVLVVMTTLCVWIFKRI
                     RNNLENSDKTNKKFSMVKLSVIFLCLGIGYAVIRRTIFIVYPPILSNGTFNFIVMQTL
                     FYLPFFILGALAFIFPHLKALFTTPSRGCTLAAALAFVAYLLNQRYGSGDAWMYETES
                     VITMVLGLWMVNVVFSFGHRLLNFQSARVTYFVNASLFIYLVHHPLTLFFGAYITPHI
                     TSNWLGFLCGLIFVVGIAIILYEIHLRIPLLKFLFSGKPVVKRENDKAPAR"
     gene            6003. .7538
                     /gene="mdoG"
                     /note="synonym: Z1683"
     CDS             6003. .7538
                     /gene="mdoG"
                     /function="enzyme; Osmotic adaptation"
                     /note="Residues 1 to 511 of 511 are 99.80 pct identical to
                     residues 1 to 511 of 511 from Escherichia coli K-12 Strain
                     MG1655: B1048"
                     /codon_start=1
                     /transl_table=11
                     /product="periplasmic glucans biosynthesis protein"
                     /protein_id="AAG55794.1"
                     /db_xref="GI:12514580"
                     /translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAQSLAGKGYEAPK
                     SNLPSVFRDMKYADYQQIQFNHDKAYWNNLKTPFKLEFYHQGMYFDTPVKINEVTATA
                     VKRIKYSPDYFTFGDVQHDKDTVKDLGFAGFKVLYPINSKDKNDEIVSMLGASYFRVI
                     GAGQVYGLSARGLAIDTALPSGEEFPRFKEFWIERPKPTDKRLTIYALLDSPRATGAY
                     KFVVMPGRDTVVDVQSKIYLRDKVGKLGVAPLTSMFLFGPNQPSPANNYRPELHDSNG
                     LSIHAGNGEWIWRPLNNPKHLAVSSFSMENPQGFGLLQRGRDFSRFEDLDDRYDLRPS
                     AWVTPKGEWGKGSVELVEIPTNDETNDNIVAYWTPDQLPEPGKEMNFKYTITFSRDED
                     KLHAPDNAWVQQTRRSTGDVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDTPVTAQTS
                     IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDAKKTTEMRAALVNADQTLSETWSYQLP
                     ANE"
     gene            7531. .10074
                     /gene="mdoH"
                     /note="synonym: Z1684"
     CDS             7531. .10074
                     /gene="mdoH"
                     /function="enzyme; Osmotic adaptation"

  Query Match           49.4%;  Score 225.4;  DB 1;  Length 10190;
  Best Local Similarity  69.7%;  Pred. No. 4.5e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       1113 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 1172

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCA---TGAAATGGCACATGCATCC 117
              || ||   | ||||   |    | |||  | || |||   ||    |   ||     |
Db       1173 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 1232

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              |||||| | |||  | |||  ||||||||||||||||   | |||| |||| ||||||||||
Db       1233 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 1292

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||   ||||| ||||| |||||    || ||||||||||   |||  ||||||||| ||
Db       1293 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 1352

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              ||||| || |||||  | ||| | ||  | || || ||||| ||   ||| ||  | ||
Db       1353 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 1412

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              |  |||||  |.|| |||||||||| |||| ||||  |||| || ||    |||| |||
Db       1413 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 1472

Qy        358 GGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAG 417
              ||    ||||  ||   |||| | ||||| |||||| |||||| ||| |  ||   ||||
Db       1473 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 1532

Qy        418 GTTGGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||| ||||| |||||||| || ||| | ||||| |||
Db       1533 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 1571


RESULT 13
AP002554
LOCUS       AP002554               327773 bp    DNA     linear   BCT 21-DEC-2002
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION   AP002554 BA000007
VERSION     AP002554.1  GI:13360491
KEYWORDS    .
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
   PUBMED   10734605
REFERENCE   2
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
   PUBMED   11108008
REFERENCE   3
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
   PUBMED   11111050
REFERENCE   4
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
   PUBMED   11258796
REFERENCE   5  (bases 1 to 327773)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .327773
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
```

/db_xref="taxon:83334"
gene 150. .2258
/gene="ECs1032"
CDS 150. .2258
/gene="ECs1032"
/note="similar to YCBY_ECOLI gi|1787181 percent identity 99 in 702 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative oxidoreductase"
/protein_id="BAB34455.1"
/db_xref="GI:13360492"
/translation="MNSLFASTARGLEELLKTELENLGAVECQVVQGGVHFKGDTRLV
YQSLMWSRLASRIMLPLGECKVYSDLDLYLGVQAINWTEMFNPGATFAVHFSGLNDTI
RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRDAPDIRVNVWLHKETASIALDLSGDGLH
LRGYRDRAGIAPIKETLAAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMLATDRAPGL
HRGRWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIQRARTNAR
LAGIGELITFEVKDVAQLTNPLPKEPYGTVLSNPPYGERLDSEPALIALHSLLGRIMK
NQFGGWNLSLFSASPDLLSCLQLRADKQYKAKNGPLDCVQKNYHVAESTPDSKPAMVA
EDYANRLRKNLKKFEKWARQEGIECYRLYDADLPEYNVAVDRYADWVVVQEYAPPKTI
DAHKARQRLFDIIAATISVLGIAPNKLVLKTRERQKGKNQYQKLGEKGEFLEVTEYNA
HLWVNLTDYLDTGLFLDHRIARRMLGQMSKGKDFLNLFSYTGSATVHAGLGGARSTTT
VDMSRTYLEWAERNLRLNGLTGRAHRLIQADCLAWLREANEQFDLIFIDPPTFSNSKR
MEDAFDVQRDHLALMKDLKRLLRAGGTIMFSNNKRGFRMDLDGLAKLGLKAQEITQKT
LSQDFARNRQIHNCWLITAA"
gene 2270. .4177
/gene="ECs1033"
CDS 2270. .4177
/gene="ECs1033"
/note="similar to UUP_ECOLI gi|1787182 percent identity 99 in 635 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative ATP-binding component of a transport system"
/protein_id="BAB34456.1"
/db_xref="GI:13360493"
/translation="MSLISMHGAWLSFSDAPLLDNAELHIEDNERVCLVGRNGAGKST
LMKILNREQGLDDGRIIYEQDLIVARLQQDPPRNVEGSVYDFVAEGIEEQAEYLKRYH
DISRLVMNDPSEKNLNELAKVQEQLDHHNLWQLENRINEVLAQLGLDPNVALSSLSGG
WLRKAALGRALVSNPRVLLLDEPTNHLDIETIDWLEGFLKTFNGTIIFISHDRSFIRN
MATRIVDLDRGKLVTYPGNYDQYLLEKEEALRVEELQNAEFDRKLAQEEVWIRQGIKA
RRTRNEGRVRALKAMRRERGERREVMGTAKMQVEEASRSGKIVFEMEDVCYQVDGKQL
VKDFSAQVLRGDKIALIGPNGCGKTTLLKLMLGQLQADSGRIHVGTKLEVAYFDQHRA
ELDPDKTVMDNLAEGKQEVMVNGKPRHVLGYLQDFLFHPKRAMTPVRALSGGERNRLL
LARLFLKPSNLLILDEPTNDLDVETLELLEELIDSYQGTVLLVSHDRQFVDNTVTECW
IFEGGGKIGRYVGGYHDARGQQEQYVALKQPAVKKNEEPAAPKAETVKRSSSKLSYKL
QRELEQLPQLLEDLEAKLEALQTQVADASFFSQPHEQTQKVLADMAAAEQELEQAFER
WEYLEALKNGG"
gene 4307. .5560
/gene="ECs1034"
CDS 4307. .5560
/gene="ECs1034"
/note="similar to PQIA_ECOLI gi|1787183 percent identity 99 in 417 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="paraquat-inducible protein A"
/protein_id="BAB34457.1"
/db_xref="GI:13360494"
/translation="MCEHHHAAKHILCSQCDMLVALPRLEHGQKAACPRCGTTLTVAW
DAPRQRPTAYALAALFMLLLSNLFPFVNMNVAGVTSEITLLEIPGVLFSEDYASLGTF
FLLFVQLVPAFCLITILLLVNRAELPVRLKEQLARVLFQLKTWGMAEIFLAGVLVSFV
KLMAYGSIGVGSSFLPWCLFCVLQLRAFQCVDRRWLWDDIAPMPELRQPLKPGVTGIR
QGLRSCSCCTAILPADEPVCPRCGTKGYVRRRNSLQWTLALLVTSIMLYLPANILPIM
VTDLLGSKMPSTILAGVILLWSEGSYPVAAVIFLASIMVPTLKMIAIAWLCWDAKGHG
KRDSERMHLIYEVVEFVGRWSMIDVFVIAVLSALVRMGGLMSIYPAMGALMFALVVIM
TMFSAMTFDPRLSWDRQPESEHEES"
gene 5565. .7205
/gene="ECs1035"
CDS 5565. .7205
/gene="ECs1035"
/note="similar to PQIB_ECOLI gi|1787184 percent identity 99 in 546 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="paraquat-inducible protein B"
/protein_id="BAB34458.1"
/db_xref="GI:13360495"
/translation="MESNNGEAKIQKVKNWSPVWIFPIVTALIGAWVLFYHYSHQGPE
VTLITANAEGIEGGKTTIKSRSVDVGVVESATLADDLTHVEIKARLNSGMEKLLHKDT
VFWVVKPQIGREGISGLGTLLSGVYIELQPGAKGSKMDKYDLLDSPPLAPPDAKGIRV
VLDSKKAGQLSPGDPVLFRGYRVGSVETSTFDTQKRNISYQLFINAPYDRLVTSNVRF
WKDSGIAVDLTSAGMRVEMGSLTTLLSGGVSFDVPEGLDLGQPVAPKTAFVLYDDQKS
IQDSLYTDHIDYLMFFKDSVRGLQPGAPVEFRGIRLGTVSKVPFFAPNMRQTFNDDYR
IPVLIRIEPERLKMQLGENADVVEHLGELLKRGLRGSLKTGNLVTGALYVDLDFYPNT
PAITGIREFNGYQIIPTVSGGLAQIQQRLMEALDKINKLPLNPMIEQATSTLSESQRT
MKNLQTTLDSMNKILASQSMQQLPTDMQSTLRELNRSMQGFQPGSAAYNKMVADMQRL
DQVLRELQPVLKTLNEKSNALVFEAKDKKDPEPKRAKQ"
gene 7217. .7765
/gene="ECs1036"
CDS 7217. .7765
/gene="ECs1036"
/note="similar to YMBA_ECOLI gi|1787185 percent identity 98 in 182 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB34459.1"
/db_xref="GI:13360496"
/translation="MTIAALWLAGCSSGEINKNYYQLPVVQSGTQSTASQGNRLLWVE
QVAVPDYLAGNGVVYQTSDVKYVIANNNLWASPLDQQLRNTLVANLSTQLPGWVVASQ
PLGSAQDTLNVTVTEFNGRYDGKVIVSGEWLLNHQGQLIKRPFRLEGVQTQDGYDEMV
KVLAGVWSQEAASIAQEIKRLP"
gene 8021. .8188
/gene="ECs1037"
CDS 8021. .8188
/gene="ECs1037"
/note="identical to RMF_ECOLI gi|1787186 (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="ribosome modulation factor"
/protein_id="BAB34460.1"
/db_xref="GI:13360497"
/translation="MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGW
REAMADRVVMA"
gene complement(8258. .8776)
/gene="ECs1038"
CDS complement(8258. .8776)
/gene="ECs1038"
/note="similar to FABA_ECOLI gi|1787187 percent identity 100 in 172 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11

Query Match 49.4%; Score 225.4; DB 1; Length 327773;
Best Local Similarity 69.7%; Pred. No. 6e-57;
Matches 320; Conservative 0; Mismatches 136; Indels 3; Gaps 1;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db 326421 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 326480

Qy 61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCA---TGAAATGGCACATGCATCC 117
|| || | |||| | | ||| | || ||| || | || |
Db 326481 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 326540

Qy 118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177

```
                ||||||  | |||  | |||  ||||||||||||||||   | |||| |||| |||||||||
Db      326541 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 326600

Qy         178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
                |||   ||||| ||||| |||||     || |||||||||   |||  ||||||||| ||
Db      326601 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 326660

Qy         238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
                ||||| || |||||  | ||| | ||  | || || ||||| ||   ||| ||  | ||
Db      326661 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 326720

Qy         298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
                |  |||||  | || ||||||||||| |||| ||||  |||| || ||     |||| |||
Db      326721 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 326780

Qy         358 GGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAG 417
                ||    ||||  ||   |||| | ||||| |||||| |||||| ||| |  ||   ||||
Db      326781 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 326840

Qy         418 GTTGGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
                ||||| ||||| |||||||| || ||| | ||||| |||
Db      326841 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 326879
```

```
RESULT 14
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
```

```
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           49.1%;  Score 224;  DB 1;  Length 4680;
  Best Local Similarity  68.2%;  Pred. No. 1.1e-56;
  Matches  311;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||   | ||||  |    | | | |   | ||     |  ||      ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | ||||||||||| |||||   ||| |||||||||  ||    |||| || || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 15
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
```

```
repeat_region   complement(<1. .48)
                /note="IS3"
promoter        <1. .5
                /note="factor Sigma70; predicted +1 start at 1094689"
promoter        5. .32
                /note="factor Sigma70; predicted +1 start at 1094716"
gene            69. .392
                /gene="b1028"
CDS             69. .392
                /gene="b1028"
                /function="orf; Unknown"
                /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                SW: P43098"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74113.1"
                /db_xref="GI:1787266"
                /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                AKVFK"
promoter        275. .299
                /gene="b1028"
                /note="factor Sigma32; predicted +1 start at 1094983"
gene            389. .1375
                /gene="ycdU"
                /note="synonym: b1029"
CDS             389. .1375
                /gene="ycdU"
                /function="orf; Unknown"
                /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                to 191 residues of an approx. 288 aa protein Y320_MYCGE
                SW: P47562"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74114.1"
                /db_xref="GI:1787267"
                /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
protein_bind    1253. .1280
                /gene="ycdU"
                /note="central position to predicted promoter: -104"
                /bound_moiety="DeoR predicted site"
protein_bind    1253. .1280
                /gene="ycdU"
                /note="No predicted promoter"
                /bound_moiety="DeoR predicted site"
promoter        1337. .1365
                /gene="ycdU"
                /note="factor Sigma70; predicted +1 start at 1096049"
gene            1494. .1745
                /gene="b1030"
CDS             1494. .1745
                /gene="b1030"
                /function="orf; Unknown"
                /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                P49065"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74115.1"
                /db_xref="GI:1787268"
                /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
gene            1513. .1926
                /gene="ycdV"
                /note="synonym: b1031"
CDS             1513. .1926
                /gene="ycdV"
                /function="putative structure; Not classified"
                /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                P05759"
                /codon_start=1
                /transl_table=11
                /product="putative ribosomal protein"
                /protein_id="AAC74116.1"
                /db_xref="GI:1787269"
                /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
gene            complement(2111. .2198)
                /gene="serX"
                /note="synonym: b1032"
tRNA            complement(2111. .2198)
                /gene="serX"
                /product="tRNA-Ser"
                /note="anticodon: GGA; CG Site No. 17878"
                /function="RNA; tRNA"
promoter        2269. .2296
                /note="factor Sigma70; predicted +1 start at 1096980"
promoter        2338. .2365
                /note="factor Sigma70; predicted +1 start at 1097049"
promoter        2355. .2383
                /note="factor Sigma70; predicted +1 start at 1097067"
gene            2393. .3370
                /gene="ycdW"
                /note="synonym: b1033"
CDS             2393. .3370
                /gene="ycdW"
                /function="putative enzyme; Not classified"
                /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                to 177 residues of an approx. 312 aa protein YPRA_CORGL
                SW: P45637"
                /codon_start=1
                /transl_table=11
                /product="putative dehydrogenase"
                /protein_id="AAC74117.1"
                /db_xref="GI:1787270"
                /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
promoter        3278. .3308
                /gene="ycdW"
                /note="factor Sigma70; predicted +1 start at 1097992"
promoter        3383. .3410
                /note="factor Sigma70; predicted +1 start at 1098094"
promoter        3396. .3423
```

```
Query Match             49.1%;  Score 224;  DB 1;  Length 10346;
Best Local Similarity   68.2%;  Pred. No. 1.2e-56;
Matches  311;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||   | ||||   |    | | | |    | ||    |  ||     ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
```

```
                |  ||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| ||||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

Search completed: March 18, 2004, 15:02:48
Job time : 2312.32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64624	Aac64624 AgfA::PT3
	2	403.2	88.4	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	403.2	88.4	456	2	AAT74142	Aat74142 Salmonell
	4	403.2	88.4	456	3	AAC64617	Aac64617 Salmonell
	5	360	78.9	456	3	AAC64626	Aac64626 AgfA::PT3
	6	353.6	77.5	456	3	AAC64625	Aac64625 AgfA::PT3
	7	352	77.2	456	3	AAC64628	Aac64628 AgfA::PT3
	8	352	77.2	456	3	AAC64622	Aac64622 AgfA::PT3
	9	352	77.2	456	3	AAC64629	Aac64629 AgfA::PT3
	10	352	77.2	456	3	AAC64623	Aac64623 AgfA::PT3
	11	348.8	76.5	456	3	AAC64630	Aac64630 AgfA::PT3
	12	348.8	76.5	456	3	AAC64627	Aac64627 AgfA::PT3
	13	345.6	75.8	456	3	AAC64631	Aac64631 AgfA::PT3
	14	284.2	62.3	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	284.2	62.3	361	2	AAT74141	Aat74141 Salmonell
	16	222.4	48.8	456	3	AAC64619	Aac64619 Escherich
	17	222.4	48.8	456	9	ACF36153	Acf36153 E. coli C
	18	210	46.1	646	2	AAQ62647	Aaq62647 Fibronect
	19	149.6	32.8	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
c	21	49	10.7	78	3	AAC64609	Aac64609 AgfA (SEF
	22	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	23	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	24	48	10.5	78	3	AAC64610	Aac64610 AgfA (SEF
	25	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	26	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
	27	41.6	9.1	72	9	ACF36148	Acf36148 E. coli C
c	28	41.2	9.0	78	3	AAC64605	Aac64605 SefA (SEF
	29	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
	30	40.6	8.9	78	3	AAC64606	Aac64606 SefA (SEF
	31	35.6	7.8	1965	4	AAS52627	Aas52627 E. coli D
	32	35.6	7.8	1965	7	ACA32662	Aca32662 Prokaryot
c	33	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	34	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
	35	34.6	7.6	456	3	AAC64620	Aac64620 Escherich
c	36	34.6	7.6	1506	6	ABQ90743	Abq90743 M. capsul
	37	34	7.5	456	3	AAC64618	Aac64618 Salmonell
c	38	34	7.5	2000	7	ADA71938	Ada71938 Rice gene
	39	34	7.5	2583	7	ACA31722	Aca31722 Prokaryot
	40	34	7.5	2886	7	ACC59886	Acc59886 Saccharom
	41	34	7.5	3300	3	AAZ55699	Aaz55699 DNA encod
	42	33.6	7.4	549	3	AAC95305	Aac95305 Cat flea
	43	33.4	7.3	3990	2	AAX61785	Aax61785 B. burgdo
	44	33.4	7.3	5100	2	AAX20271	Aax20271 Borrelia
	45	33.4	7.3	7766	2	AAT85043	Aat85043 Borrelia

ALIGNMENTS

RESULT 1
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.5e-141;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.4%; Score 403.2; DB 2; Length 456;
Best Local Similarity 92.8%; Pred. No. 2.9e-123;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||  | | | |  |  | |    |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             88.4%;  Score 403.2;  DB 2;  Length 456;
  Best Local Similarity   92.8%;  Pred. No. 2.9e-123;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |  | |      |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.4%; Score 403.2; DB 3; Length 456;
Best Local Similarity 92.8%; Pred. No. 2.9e-123;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | | |    ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 78.9%; Score 360; DB 3; Length 456;
Best Local Similarity 86.8%; Pred. No. 6.4e-109;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | | |    ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||| |||   |   ||||||   | |    | |   ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 6
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match             77.5%;  Score 353.6;  DB 3;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 8.5e-107;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||        |   ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 7
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             77.2%;  Score 352;  DB 3;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.9e-106;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||     ||  ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
               || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             77.2%;  Score 352;  DB 3;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.9e-106;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             77.2%;  Score 352;  DB 3;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.9e-106;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |   |  | |    |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||      |           ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.9e-106;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||    | | | |  | |     |  ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||        |  || | ||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match            76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 3.3e-105;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||   | | | |  |   | |     |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |· |||  |     |  |  ||  |   |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 3.3e-105;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||   | | | |  |   | |     |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||  |  |      | || |   || |  |  | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36355.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 75.8%; Score 345.6; DB 3; Length 456;
Best Local Similarity 84.9%; Pred. No. 3.8e-104;
Matches 387; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 14
AAQ73066
ID AAQ73066 standard; DNA; 361 BP.
XX
AC AAQ73066;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella enteritidis.
XX
FH Key Location/Qualifiers
FT CDS 1. .359
FT /*tag= a
FT /note= "AgfA"
FT misc_feature 37. .60
FT /*tag= d
FT /note= "TAF5 primer (pair with TAF6)"
FT misc_feature 52. .69
FT /*tag= b
FT /note= "TAF3 primer (pair with TAF4)"
FT misc_feature complement(103. .129)
FT /*tag= e
FT /note= "TAF6 primer (pair with TAF5)"
FT misc_feature complement(292. .402)
FT /*tag= c
FT /note= "TAF4 primer (pair with TAF3)"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             62.3%;  Score 284.2;  DB 2;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 8.4e-84;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             62.3%;  Score 284.2;  DB 2;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 8.4e-84;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 18, 2004, 09:11:57
Job time : 289.405 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	403.2	88.4	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	284.2	62.3	361	1	US-08-233-788A-56	Sequence 56, Appl
	3	33.4	7.3	7766	4	US-09-125-619-3	Sequence 3, Appli
c	4	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	5	31.2	6.8	2436	4	US-09-540-236-492	Sequence 492, App
c	6	31.2	6.8	65792	4	US-09-596-002-31	Sequence 31, Appl
	7	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
	8	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	9	30.2	6.6	1155	4	US-09-328-352-1775	Sequence 1775, Ap
c	10	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
	11	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	12	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, App
c	13	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, App
c	14	29.4	6.4	3494	4	US-09-023-655-1006	Sequence 1006, Ap
c	15	29.2	6.4	2052	1	US-08-443-104-5	Sequence 5, Appli
c	16	29.2	6.4	2052	1	US-08-238-130-6	Sequence 6, Appli
c	17	29.2	6.4	2052	1	US-08-442-859-5	Sequence 5, Appli
c	18	29.2	6.4	2052	2	US-08-398-489-5	Sequence 5, Appli
c	19	29.2	6.4	2052	2	US-08-894-772-1	Sequence 1, Appli
c	20	29.2	6.4	2052	2	US-09-207-844-1	Sequence 1, Appli
c	21	29.2	6.4	2052	4	US-09-252-509-1	Sequence 1, Appli
c	22	29.2	6.4	2052	5	PCT-US95-05534-5	Sequence 5, Appli
c	23	29	6.4	13794	4	US-08-956-171E-54	Sequence 54, Appl
c	24	29	6.4	1830121	4	US-09-557-884-1	Sequence 1, Appli
c	25	29	6.4	1830121	4	US-09-643-990A-1	Sequence 1, Appli
	26	28.8	6.3	720	4	US-09-328-352-354	Sequence 354, App
	27	28.8	6.3	1662	1	US-08-565-386-2	Sequence 2, Appli
	28	28.8	6.3	1836	4	US-09-252-991A-15831	Sequence 15831, A
	29	28.8	6.3	2055	4	US-09-252-991A-15771	Sequence 15771, A
c	30	28.8	6.3	2106	4	US-09-252-991A-15891	Sequence 15891, A
	31	28.8	6.3	4529	1	US-08-565-386-1	Sequence 1, Appli
c	32	28.8	6.3	8906	4	US-09-027-169-5	Sequence 5, Appli
c	33	28.4	6.2	1520	4	US-09-484-970B-30	Sequence 30, Appl
	34	28.2	6.2	2116	3	US-09-000-041A-1	Sequence 1, Appli
	35	28.2	6.2	2426	3	US-09-028-327-2	Sequence 2, Appli
	36	28.2	6.2	2426	4	US-09-571-078A-2	Sequence 2, Appli
	37	28.2	6.2	2812	1	US-08-920-812-16	Sequence 16, Appl
	38	28.2	6.2	2812	1	US-08-920-827-16	Sequence 16, Appl
	39	28.2	6.2	2812	1	US-08-921-177-16	Sequence 16, Appl
	40	28.2	6.2	2812	1	US-08-362-577C-16	Sequence 16, Appl
	41	28.2	6.2	2812	2	US-08-920-828-16	Sequence 16, Appl
c	42	28.2	6.2	4973	3	US-09-381-862-6	Sequence 6, Appli
	43	28.2	6.2	8078	4	US-09-702-251-3	Sequence 3, Appli
	44	28	6.1	906	4	US-09-134-001C-383	Sequence 383, App
c	45	28	6.1	1665	4	US-09-540-236-1858	Sequence 1858, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 88.4%; Score 403.2; DB 1; Length 456;
Best Local Similarity 92.8%; Pred. No. 5e-135;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||   ||     | | | |   | |     |   ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             62.3%;  Score 284.2;  DB 1;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 2.9e-92;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

```
RESULT 3
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-09-125-619-3

  Query Match              7.3%;  Score 33.4;  DB 4;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 0.33;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||      ||  ||| |  | | |  | | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |   |   | ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551
```

```
Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |  || ||| |   |    | ||| |           | ||| ||||| |     |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
               || ||||       | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 4
5486473-3/c
;Patent No. 5486473
;    APPLICANT: FUKITA, HIROYUKI;YOSHIDA, IWAO;TAKAGI, MITSUO;
;MANABE, SADAO;FUKAI, KONOSUKE
;    TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
;    NUMBER OF SEQUENCES: 16
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,049
;      FILING DATE: 09-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 809,255
;      FILING DATE: 18-DEC-1991
;      APPLICATION NUMBER: 279,685
;      FILING DATE: 05-DEC-1988
;      APPLICATION NUMBER: 932,419
;      FILING DATE: 19-NOV-1986
;SEQ ID NO:3:
;      LENGTH: 1491
5486473-3

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 1491;
  Best Local Similarity   50.0%;  Pred. No. 0.77;
  Matches   78;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              |||  |  || ||| |  |  | |||    | | | || | | |  | | |  | ||  |
Db        204 CAAGTCACTGACCGAAGCTAGGTAACAGTAACTGCGCACATCTGCGAGGTTGGCTGCTTC 145

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
               ||||    ||  ||     ||| ||    |||   |    ||  |||||  || || |
Db        144 CATGTTCATCATTTTGACATCAATGGTTGGCTTGTCTTTTGACATTATGGTCACACAACT 85

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCC 333
              |  |||  |    |||||  |||  |    | ||||
Db         84 ATCGCCTTCCAGTACCAGATCAACCCATGTAGCTCC 49


RESULT 5
US-09-540-236-492/c
; Sequence 492, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2005-001
;  CURRENT APPLICATION NUMBER: US/09/540,236
;  CURRENT FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 492
;   LENGTH: 2436
;   TYPE: DNA
;   ORGANISM: M.catarrhalis
US-09-540-236-492

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 2436;
  Best Local Similarity   53.7%;  Pred. No. 1;
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |||||||  |    | ||  |      | | |  | || || ||||| |||    | |||
Db       1511 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 1452

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |      | |  || ||||  |||| |||||| |||| |||| ||  || |    |
Db       1451 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 1392

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |  || ||  ||  |||| ||  ||||   |  |||
Db       1391 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 1348


RESULT 6
US-09-596-002-31/c
; Sequence 31, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:      Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:      PERL Program
; SEQ ID NO 31
;   LENGTH: 65792
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte template ID No. 6632636 31
;  PUBLICATION INFORMATION:
US-09-596-002-31

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 65792;
  Best Local Similarity   53.7%;  Pred. No. 7.1;
  Matches   88;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        209 TTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTA 268
              |||||||  |    | ||  |      | | |  | || || ||||| |||    | |||
Db       6819 TTACCCACATTCATCATAATCTTAAAGGCAAGTTCGGGCATGGTGCTGATTGAATTACGC 6760

Qy        269 TTGA---ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              |  |      | |  || ||||  |||| |||||| ||||  |||| ||  || |    |
Db       6759 TGAACTTCATAATCCAAAATGCCTTCATAAATAAAGCCAGTATCGCCCTCTGCACAAGAA 6700

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |      |  || ||  ||  |||| ||  ||||   |  |||
Db       6699 ACTGTCACCTCATGACCATCAACCAAAACTTCGGTTGCATTGCC 6656


RESULT 7
US-08-508-761B-1
; Sequence 1, Application US/08508761B
; Patent No. 6027920
;  GENERAL INFORMATION:
;    APPLICANT:  Joliff, Gwennael
;    APPLICANT:  Guyonvarch, Armel
;    APPLICANT:  Purification, Relano
;    APPLICANT:  Duchiron, Francis
;    APPLICANT:  Renaud, Michel
;    TITLE OF INVENTION:  System for Protein Expression and
;    TITLE OF INVENTION:  Secretion Especially in Corynebacteria
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jacobson, Price, Holman & Stern, PLLC
;      STREET:  400 Seventh St. N.W.
;      CITY:  Washington D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/508,761B
;       FILING DATE:  31-JUL-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 91/09652
;       FILING DATE:  29-JUL-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 91/09870
;       FILING DATE:  02-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Player, William E.
;       REGISTRATION NUMBER:  31,409
;       REFERENCE/DOCKET NUMBER:  P58525NA
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 638-6666
;       TELEFAX:  (202) 393-5350
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2547 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Corynebacterium Melassecola
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  420..2390
US-08-508-761B-1

  Query Match             6.8%;  Score 30.8;  DB 3;  Length 2547;
  Best Local Similarity   58.9%;  Pred. No. 1.5;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        280 CAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATT 339
              |||| |||||  |  | | |  ||| ||| || ||||   ||||  | ||   ||  |||
Db        876 CAGAGTGGTTGGACCATTGAGACCAACATTGAGCAGTACTACGCCGATAAGAACGCCATT 935

Qy        340 ACTGTCGGCCAATACGGCGGTAATAACGCC 369
              |||   || |  ||| ||  | | | ||
Db        936 GTTGTGCTCCCAATCGGTGGCGAGAGCTCC 965


RESULT 8
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
```

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             6.8%;  Score 30.8;  DB 4;  Length 1664976;
  Best Local Similarity   51.4%;  Pred. No. 60;
  Matches   71;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        288 TTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGG 347
              ||| | ||||  || ||  || | | |    ||||  || ||||    ||||| | ||
Db     360315 TTTTATAAATGTTGTCATGATGGTCAAAATCAACGAAAATAACTGTTTTATTATTCTCAT 360374

Qy        348 CCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAAT 407
              | |    |     || |||   ||  |  |  ||  || | | | | |   ||    ||
Db     360375 CAACGGTGAAAACAAGAACAAAGCTTTTATCGATATGAACTCTTTTAAAATCATTTAAAG 360434

Qy        408 GGTGCGTCAGGTTGGTTT 425
              | ||  | |||||  | |
Db     360435 GATGTCTTAGGTTTTTAT 360452


RESULT 9
US-09-328-352-1775
; Sequence 1775, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1775
;   LENGTH: 1155
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-1775

  Query Match             6.6%;  Score 30.2;  DB 4;  Length 1155;
  Best Local Similarity   51.9%;  Pred. No. 1.5;
  Matches   68;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         79 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGC 138
              | ||  || |    ||  ||  || ||||  ||| ||  |||   |||  ||| | |
Db        799 GATATTCGAGCAAATAATCAGCAATTGGCTTATGGATATGGCTATGACCAAACTTGGGTA 858

Qy        139 ATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCT 198
              ||| ||||  |   |  |  ||        |||||           | || ||  ||||||
Db        859 ATTAATCAAAAGTCTCAAGGGAAACTCAATCTTGCAGCTATTGTGGTTGACCCAAAATCT 918

Qy        199 GAAACGACCAT 209
              ||  ||||||
Db        919 AAACGGACCAT 929


RESULT 10
US-09-328-352-2055/c
; Sequence 2055, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2055
```

```
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-2055

  Query Match              6.5%;  Score 29.6;  DB 4;  Length 972;
  Best Local Similarity   61.8%;  Pred. No. 2.2;
  Matches   47;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || |||| | | | | ||| |   ||  ||||| |||| |   | | |||| | | ||
Db        558 ATTAGTGAAGTAAATAATAATTGCACCAATAATAACGCCAACTGAGCCTAGTGCATCACT 499

Qy        325 AAAAACTCCGATATTA 340
              || ||| | | || |
Db        498 CAATACTTCTAGATAA 483


RESULT 11
US-08-961-527-310
; Sequence 310, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  310:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1939 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-310

  Query Match              6.4%;  Score 29.4;  DB 4;  Length 1939;
  Best Local Similarity   56.8%;  Pred. No. 4;
  Matches   54;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy        235 GCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGA 294
              ||| |||||| |  || | |||| ||      || |||| |  || ||| ||    |||
Db        618 GCCAATGTAGTCACGGTTACGGAAAAAGACTTTACTGAATTTCCTAAGAGTGACACCAGT 677

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAA 329
              ||    | | | | | ||     | |  ||| ||||
Db        678 AAGGTTCCTATCCTAGATAGAAGTACTGCTGAAAA 712
```

```
RESULT 12
US-09-139-802-200/c
; Sequence 200, Application US/09139802
; Patent No. 6180084
; GENERAL INFORMATION:
;  APPLICANT: Ruoslahti, Erkki
;  APPLICANT: Pasqualini, Renata
;  TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing
;  TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using
;  TITLE OF INVENTION: Same
;  FILE REFERENCE: P-LJ 3203
;  CURRENT APPLICATION NUMBER: US/09/139,802
;  CURRENT FILING DATE: 1998-08-25
;  EARLIER APPLICATION NUMBER: 08/926,914
;  EARLIER FILING DATE: 1997-09-10
;  EARLIER APPLICATION NUMBER: 08/710,067
;  EARLIER FILING DATE: 1996-09-10
;  NUMBER OF SEQ ID NOS: 226
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 200
;   LENGTH: 3494
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (121)..(3024)
US-09-139-802-200

  Query Match              6.4%;  Score 29.4;  DB 3;  Length 3494;
  Best Local Similarity   45.2%;  Pred. No. 5.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   |  | || || |   |||| || ||   || |   ||||
Db       2531 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 2472

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              |  ||   ||    |    | || ||||   | | |    |       ||||
Db       2471 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 2412

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |   | | |||   | |||  |||   ||| |  |||||  |||||       |
Db       2411 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 2352

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |    | | ||  ||    ||  |   |  ||  | |    | | |||  |
Db       2351 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 2293


RESULT 13
US-09-659-786-200/c
; Sequence 200, Application US/09659786
; Patent No. 6491894
; GENERAL INFORMATION:
;  APPLICANT:  Ruoslahti, Erkki
;  APPLICANT:  Pasqualini, Renata
;  TITLE OF INVENTION:  NGR Receptor and Methods of Identifying Tumor Homing
;  TITLE OF INVENTION:  Molecules That Home to Angiogenic Vasculature Using
;  TITLE OF INVENTION:  Same
;  FILE REFERENCE:  P-LJ 3203
;  CURRENT APPLICATION NUMBER: US/09/659,786
;  CURRENT FILING DATE:  2000-09-11
;  PRIOR APPLICATION NUMBER:  08/926,914
;  PRIOR FILING DATE:  1997-09-10
;  PRIOR APPLICATION NUMBER:  08/710,067
;  PRIOR FILING DATE:  1996-09-10
;  NUMBER OF SEQ ID NOS:  226
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 200
;   LENGTH: 3494
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (121)..(3024)
US-09-659-786-200

  Query Match             6.4%;  Score 29.4;  DB 4;  Length 3494;
  Best Local Similarity   45.2%;  Pred. No. 5.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   | | || || |   |||| || ||   || |   ||||
Db       2531 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 2472

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              |  ||   ||     |     | || ||||    | | |     |         ||||
Db       2471 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 2412

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  ||||| |||||      |
Db       2411 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 2352

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||    ||  |   |  ||  | |    |  | ||| |
Db       2351 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 2293


RESULT 14
US-09-023-655-1006/c
; Sequence 1006, Application US/09023655
; Patent No. 6607879
;  GENERAL INFORMATION:
;    APPLICANT:  Cocks, Benjamin G.
;    APPLICANT:  Susan G. Stuart
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  1508
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,655
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  1006:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3494 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GENBANK
;      CLONE:  g178535
US-09-023-655-1006

  Query Match             6.4%;  Score 29.4;  DB 4;  Length 3494;
  Best Local Similarity   45.2%;  Pred. No. 5.6;
  Matches  108;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
              ||| | ||| |   | |   | | || || |   |||| || ||   || |   ||||
Db       2531 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 2472

Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              |  ||   ||     |     | || ||||    | | |     |         ||||
Db       2471 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 2412

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  ||||| |||||      |
Db       2411 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 2352

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |     | | ||  ||    ||  |   |  ||  | |    |  | ||| |
Db       2351 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 2293


RESULT 15
US-08-443-104-5/c
; Sequence 5, Application US/08443104
; Patent No. 5691162
;  GENERAL INFORMATION:
;    APPLICANT:  Shuster, Jeffrey R.
;    APPLICANT:  Madden, Mark
;    APPLICANT:  Moyer, Donna L.
;    APPLICANT:  Fuglsang, Claus
;    APPLICANT:  Branner, Sven
;    TITLE OF INVENTION:  NOVEL METALLOPROTEASE HAVING INCREASED
;    TITLE OF INVENTION:  ACTIVITY
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5691162o No. 5691162disk of No. 5691162th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Tape
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/443,104
;      FILING DATE:  17-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/398,489
;      FILING DATE:  03=MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4180.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2052 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-443-104-5

  Query Match             6.4%;  Score 29.2;  DB 1;  Length 2052;
  Best Local Similarity   49.4%;  Pred. No. 4.9;
  Matches   76;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;
```

```
Qy        300 TGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGG 359
              || || | || ||  |  ||     || | |  || | |   ||   ||    |||| |
Db        563 TGTCAGCCTCAACGGGGAGAGAAAGAACATCAACGGTGTCCTTGAGAGCGTCGACGGGGT 504

Qy        360 TAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGT 419
              |  || | |||| | |||  |||||   |  | ||  ||||   |    |   ||
Db        503 CTTTCTCGTCACGCTTGGTAAGAGGACCGGGAATCTTGCCCTCGTAGAAGCTGTTTCCGT 444

Qy        420 TGGTTTTGGCAACAACGCCACGGCTAACCAGTAT 453
              ||    | |  |  || ||  || || | ||||
Db        443 AGGAGAAGACCTCGCCGTCAGCGCCAATCTGTAT 410
```

Search completed: March 18, 2004, 20:20:36
Job time : 57.4971 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	35.6	7.8	1965	9	US-09-815-242-6264	Sequence 6264, Ap
	2	35.6	7.8	1965	12	US-10-282-122A-20532	Sequence 20532, A
	3	34	7.5	2583	12	US-10-282-122A-19592	Sequence 19592, A
	4	34	7.5	2886	9	US-09-801-368-131	Sequence 131, App
	5	33.6	7.4	549	10	US-09-991-936-1800	Sequence 1800, Ap
	6	33.4	7.3	7766	12	US-10-222-566-3	Sequence 3, Appli
	7	33.4	7.3	7766	14	US-10-222-162-3	Sequence 3, Appli
	8	33.4	7.3	7766	14	US-10-143-024-3	Sequence 3, Appli
	9	32.8	7.2	1059	9	US-09-738-626-482	Sequence 482, App
c	10	32.8	7.2	1113	9	US-09-738-626-481	Sequence 481, App
c	11	32.8	7.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
	12	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
	13	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
c	14	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
c	15	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
	16	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
	17	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
	18	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c	19	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
c	20	31.2	6.8	2394	12	US-10-282-122A-26837	Sequence 26837, A
	21	31.2	6.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	22	31	6.8	525	10	US-09-918-995-32144	Sequence 32144, A
c	23	31	6.8	616	12	US-10-424-599-88945	Sequence 88945, A
	24	31	6.8	2751	15	US-10-398-221-3563	Sequence 3563, Ap
	25	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
c	26	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
	27	30.8	6.8	1665	15	US-10-369-493-43454	Sequence 43454, A
	28	30.8	6.8	2050	14	US-10-086-510-1	Sequence 1, Appli
	29	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appli
	30	30.6	6.7	2301	9	US-09-815-242-4844	Sequence 4844, Ap
	31	30.6	6.7	2301	9	US-09-815-242-9042	Sequence 9042, Ap
c	32	30.4	6.7	392	12	US-10-282-122A-25231	Sequence 25231, A
c	33	30.4	6.7	25220	11	US-09-984-429-364	Sequence 364, App
	34	30.2	6.6	379	12	US-10-424-599-102524	Sequence 102524,
	35	30.2	6.6	3203	12	US-10-282-122A-24039	Sequence 24039, A
	36	30.2	6.6	3222	12	US-10-282-122A-6780	Sequence 6780, Ap
	37	30.2	6.6	3222	14	US-10-210-115-19	Sequence 19, Appl
	38	30.2	6.6	3222	15	US-10-369-493-24379	Sequence 24379, A
c	39	30.2	6.6	1601042	15	US-10-027-632-59064	Sequence 59064, A
c	40	30	6.6	414	14	US-10-156-761-2947	Sequence 2947, Ap
c	41	30	6.6	1074	12	US-10-282-122A-30822	Sequence 30822, A
	42	29.8	6.5	436	12	US-10-424-599-73842	Sequence 73842, A
	43	29.8	6.5	815	12	US-10-424-599-72759	Sequence 72759, A
c	44	29.8	6.5	1394	14	US-10-283-423-13	Sequence 13, Appl
c	45	29.8	6.5	1394	14	US-10-213-821-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-6264
; Sequence 6264, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6264
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Escherichia coli

```
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1965)
US-09-815-242-6264

  Query Match             7.8%;  Score 35.6;  DB 9;  Length 1965;
  Best Local Similarity   60.2%;  Pred. No. 0.27;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATG 100
              || ||| |||| |||    | |||||    |||| || |||| |||||  ||||  | | ||
Db        350 TTGCTGACAGTTCTCGTCCAGGCGTAGGTGATGAGTTGGTCACCCGAATTGTACCGCTTG 409

Qy        101 AAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGC 138
              | |  |  | |||    | || || | |   | |  ||
Db        410 AGAACGTTCCTGCTCGTGACCTGGCCCCCCTGCTCCGC 447


RESULT 2
US-10-282-122A-20532
; Sequence 20532, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20532
;   LENGTH: 1965
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-282-122A-20532

  Query Match             7.8%;  Score 35.6;  DB 12;  Length 1965;
  Best Local Similarity   60.2%;  Pred. No. 0.27;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATG 100
              || ||| |||| |||    | |||||    |||| || |||| |||||  ||||  | | ||
Db        350 TTGCTGACAGTTCTCGTCCAGGCGTAGGTGATGAGTTGGTCACCCGAATTGTACCGCTTG 409

Qy        101 AAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGC 138
              | |  |  | |||    | || || | |   | |  ||
Db        410 AGAACGTTCCTGCTCGTGACCTGGCCCCCCTGCTCCGC 447


RESULT 3
US-10-282-122A-19592
; Sequence 19592, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19592
;   LENGTH: 2583
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-10-282-122A-19592

  Query Match             7.5%;  Score 34;  DB 12;  Length 2583;
  Best Local Similarity   48.9%;  Pred. No. 1.1;
  Matches   91;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy         68 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGCCCGGACT 127
              ||| ||| | |  |  ||||   | | ||  || |  |||   ||      |  | |||
Db          1 ATGCTCAACCGAATTACCGTTCAGCTCCCGGTGGAGGGGCTTCTGTTCTGGAAACTGACG 60

Qy        128 CAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATG 187
              ||    ||  |   |    || || |||| ||||||  | ||||| |||  |    |
Db         61 GGCCGCGAGGCGATGTCCGAGTCGTTTGCGCTGACGCTGACCGTGCTCGGCACGGACGCG 120

Qy        188 CCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCC 247
              | | | | |  |    |  |   |    |||| | ||| ||    ||||  | ||  |
Db        121 CGCATCGACCGCAGTAAGCTGCTGGGCCAGCCGGTCACGGTGACCATCCCGACGCAGAGC 180
```

```
Qy        248 AGGGTG 253
              | ||
Db        181 CTGCTG 186
```

RESULT 4
US-09-801-368-131
; Sequence 131, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 131
; LENGTH: 2886
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-131

Query Match 7.5%; Score 34; DB 9; Length 2886;
Best Local Similarity 50.6%; Pred. No. 1.2;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              | ||  | | | | || | | | |||||    | ||    || |   |    |||| | |
Db       1501 AACAACAACAATAACGATAACGATAACGATAACAATAATAGCAATAATAATAATAACAAT 1560

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | ||| | |    | | | |||  |  ||  ||||||  ||  | | || |    |||  |
Db       1561 AATAATAATAACAATAATAATAATAACAATAATAATAACAATAACAACAACAACAACAAT 1620

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
              || ||     ||| || |      || |   || | |||||
Db       1621 AATAATAATAATAGTAATAATATTAAAAATAACGTTGATAAC 1662
```

RESULT 5
US-09-991-936-1800
; Sequence 1800, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisnewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1800
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-1800

Query Match 7.4%; Score 33.6; DB 10; Length 549;
Best Local Similarity 61.4%; Pred. No. 0.67;
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy         41 TTTCTGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATG 100
              || ||| |||| |||   | |||||   |||| || |||| |||||  ||||  | | ||
Db        355 TTGCTGACAGTTCTCGTCCAGGCGTAGGTGATGAGTTGGTCACCCGAATTGTACCGCTTG 414

Qy        101 AAATGGCACATGCATCCGGCCCGGACTC 128
              | |  |  | |||    | || |  | |
Db        415 AGAACGTTCCTGCTCGTGACCTGCCCCC 442
```

RESULT 6
US-10-222-566-3
; Sequence 3, Application US/10222566
; Publication No. US20040044192A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234USD3
; CURRENT APPLICATION NUMBER: US/10/222,566
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/125,619
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-222-566-3

Query Match 7.3%; Score 33.4; DB 12; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||    |    ||||     ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |    |  | |||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |   | ||| |          | |||  |||||  |   |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||       | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642
```

```
RESULT 7
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD4
;  CURRENT APPLICATION NUMBER: US/10/222,162
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-222-162-3

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |    ||||      ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |    |  ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    | ||| |           | |||  |||||  |    |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||        | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 8
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
;  APPLICANT: NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE: UTSH:234USD1
;  CURRENT APPLICATION NUMBER: US/10/143,024
;  CURRENT FILING DATE:  2002-05-10
;  PRIOR APPLICATION NUMBER: 09/125,619
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-10-143-024-3

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 7766;
  Best Local Similarity   47.4%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  ||||||  ||||     |    ||||      ||  ||| |  | | |  |  | |
Db       4432 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 4491

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              | | |  || ||  ||    | |    |  ||||||||||  | |||    |||  ||
Db       4492 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 4551

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    | ||| |           | |||  |||||  |    |
Db       4552 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 4611

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
              || ||||        | |  ||| | || |
Db       4612 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 4642


RESULT 9
US-09-738-626-482
; Sequence 482, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 482
;   LENGTH: 1059
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-482

  Query Match             7.2%;  Score 32.8;  DB 9;  Length 1059;
  Best Local Similarity   47.2%;  Pred. No. 1.8;
  Matches  100;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | | |     | | |  |||||  ||     |||  ||  | |||||   |||||| |
Db         76 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 135

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |     | | | ||  | |     || ||  ||     |      |||
Db        136 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 195

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
```

```
                | || | |   ||  | |     ||| | ||||||   || |     ||| ||| | |
Db        196 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 255

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              | | ||  |||  |  ||  |  | ||||
Db        256 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 287
```

RESULT 10
US-09-738-626-481/c
; Sequence 481, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 481
; LENGTH: 1113
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-481

Query Match 7.2%; Score 32.8; DB 9; Length 1113;
Best Local Similarity 47.2%; Pred. No. 1.9;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | |  |     | | |  |||||  ||    |||  ||  | |||||   |||||| |
Db        972 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 913

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |    |  | |  || | |     || ||  ||     |      |||
Db        912 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 853

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              | || | |   ||  | |     ||| | ||||||   || |     ||| ||| | |
Db        852 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 793

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              | | ||  |||  |  ||  |  | ||||
Db        792 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 761
```

RESULT 11
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 7.2%; Score 32.8; DB 9; Length 3309400;
Best Local Similarity 47.2%; Pred. No. 1.3e+02;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              || | |  |     | | |  |||||  ||    |||  ||  | |||||   |||||| |
Db     451820 GAGGACGTTTCCACCGCAGAGACCACCACTGCAAGCTCTTCCGCTAACGCATCCGATGCA 451761

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              | |   |    |  | |  || | |     || ||  ||     |      |||
Db     451760 GCCGGTGAAAAAGTAACCATCACCGTCTACACCTCTGAGCCTGAGGAAAAGGTCGATGAG 451701

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              | || | |   ||  | |     ||| | ||||||   || |     ||| ||| | |
Db     451700 ATCAACAAGGCGTTCATGGAAGCCAACCCAGATATTGAGGTTGAGGTGTACCGCGCTGGT 451641

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGA 395
              | | ||  |||  |  ||  |  | ||||
Db     451640 ACTGGCGATCTGACTGCTCGCATTGAAGCTGA 451609
```

RESULT 12
US-10-333-631-1
; Sequence 1, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
; APPLICANT: Bayer AG
; TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
; TITLE OF INVENTION: ENZYME
; FILE REFERENCE: LIO121 Foreign Countries
; CURRENT APPLICATION NUMBER: US/10/333,631
; CURRENT FILING DATE: 2003-01-23
; PRIOR APPLICATION NUMBER: US 60/221,917
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/280,112
; PRIOR FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (90)..(90)
; OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

Query Match 7.1%; Score 32.2; DB 14; Length 568;
Best Local Similarity 52.6%; Pred. No. 2.1;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCC 214
              | ||| |    |  ||||| |||||   ||   || ||    ||||| | |  |  ||
Db        103 CTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCTGAAATGGTCCGTCACC 162

Qy        215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              ||  |    || || | || | || | ||| ||||||||    | |  ||   ||
Db        163 CAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTACCTGGATCTGCGGATA 222

Qy        275 TGACTCAGAATGG 287
              | ||   |||
Db        223 GCCCACATGCTGG 235


RESULT 13
US-10-333-631-4
; Sequence 4, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer AG
;  TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
;  TITLE OF INVENTION:  ENZYME
;  FILE REFERENCE: LIO121 Foreign Countries
;  CURRENT APPLICATION NUMBER: US/10/333,631
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US 60/221,917
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/280,112
;  PRIOR FILING DATE: 2001-04-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-333-631-4

  Query Match             7.1%;  Score 32.2;  DB 14;  Length 972;
  Best Local Similarity   52.6%;  Pred. No. 2.8;
  Matches   70;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCC 214
              | ||| |    |  ||||| |||||   ||   || ||    ||||| | |  |  ||
Db        202 CTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCTGAAATGGTCCGTCACC 261

Qy        215 AGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              ||  |    || || | || | || | ||| ||||||||    | |  ||   ||
Db        262 CAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTACCTGGACCTGCGGATA 321

Qy        275 TGACTCAGAATGG 287
              | ||   |||
Db        322 GCCCACATGCTGG 334


RESULT 14
US-10-369-493-33133/c
; Sequence 33133, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33133
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Xylella fastidiosa
US-10-369-493-33133

  Query Match             7.1%;  Score 32.2;  DB 15;  Length 1083;
  Best Local Similarity   59.1%;  Pred. No. 3;
  Matches   55;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTG 421
              ||  | ||| |||||   |  | |||||     ||  || | ||| | ||| ||||
Db        488 ATCTCACCGAGCTGGGATAAAAAACCGCGCACAATCACACCATAACGCTGCTTCAGCCAC 429

Qy        422 GTTTTGGCAACAACGCCACGGCTAACCAGTATT 454
              |||| ||||| || |||    | ||  | |||
Db        428 TTTTTAGCAACCACCCCAGCAGTCACACGCATT 396


RESULT 15
US-10-282-122A-27869/c
; Sequence 27869, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27869
;   LENGTH: 4830
;   TYPE: DNA
;   ORGANISM: Mycobacterium leprae
US-10-282-122A-27869

  Query Match             7.1%;  Score 32.2;  DB 12;  Length 4830;
  Best Local Similarity   51.0%;  Pred. No. 6.7;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        224 ATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGA 283
```

```
            ||| ||     | |||  | | | |   |||      |    | |  ||| |  |    ||
Db       4229 ATGATAGTAAAGACGACTTCGCCGACATTGCCTCCGACCAGAATCGTGACCGCATCACGA 4170

Qy        284 ATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTG 343
              |  |  |  | ||  |||  || |||||||| | | ||   |  || || | ||    |
Db       4169 ACGCCTGCCCACATGCTGCGGCCTTCGACCAGGGCATCGAGCAGTACACCCAGATCGTCG 4110

Qy        344 TCGGCCAATACGGCGGTAATAACGCCGCG 372
              |||| ||| |||  |  |    | || ||
Db       4109 TCGGTCAAGACGATGTCAGCGGCTCCTCG 4081
```

Search completed: March 19, 2004, 03:49:40
Job time : 261.734 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-15	Sequence 15, Appl
2	403.2	88.4	456	6	US-08-233-642A-56	Sequence 56, Appl
3	403.2	88.4	456	23	US-09-543-407-1	Sequence 1, Appli
4	360	78.9	456	23	US-09-543-407-19	Sequence 19, Appl
5	353.6	77.5	456	23	US-09-543-407-17	Sequence 17, Appl
6	352	77.2	456	23	US-09-543-407-11	Sequence 11, Appl
7	352	77.2	456	23	US-09-543-407-13	Sequence 13, Appl
8	352	77.2	456	23	US-09-543-407-23	Sequence 23, Appl
9	352	77.2	456	23	US-09-543-407-25	Sequence 25, Appl
10	348.8	76.5	456	23	US-09-543-407-21	Sequence 21, Appl
11	348.8	76.5	456	23	US-09-543-407-27	Sequence 27, Appl
12	345.6	75.8	456	23	US-09-543-407-29	Sequence 29, Appl
13	284.2	62.3	361	6	US-08-233-642A-54	Sequence 54, Appl
14	222.4	48.8	456	23	US-09-543-407-3	Sequence 3, Appli
15	222.4	48.8	477	17	US-09-252-691-172	Sequence 172, App
16	222.4	48.8	477	17	US-09-252-691C-172	Sequence 172, App
17	222.4	48.8	477	51	US-10-417-886-172	Sequence 172, App
18	222.4	48.8	675	13	US-08-978-878-3	Sequence 3, Appli
19	222.4	48.8	675	31	US-09-741-873B-3	Sequence 3, Appli
20	220	48.2	453	92	US-60-352-946-1	Sequence 1, Appli
21	220	48.2	453	101	US-60-444-371-1	Sequence 1, Appli
22	180	39.5	396	13	US-08-978-878-1	Sequence 1, Appli
23	155.2	34.0	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
C 28	49	10.7	78	23	US-09-543-407-52	Sequence 52, Appl
29	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
30	48	10.5	78	23	US-09-543-407-53	Sequence 53, Appl
31	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
C 32	41.2	9.0	78	23	US-09-543-407-48	Sequence 48, Appl
33	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
34	40.6	8.9	78	23	US-09-543-407-49	Sequence 49, Appl
35	35.8	7.9	561	17	US-09-252-691-171	Sequence 171, App
36	35.8	7.9	561	17	US-09-252-691C-171	Sequence 171, App
37	35.8	7.9	561	51	US-10-417-886-171	Sequence 171, App
38	35.6	7.8	1965	1	PCT-US02-03987-6264	Sequence 6264, Ap
39	35.6	7.8	1965	33	US-09-815-242-6264	Sequence 6264, Ap
40	35.6	7.8	1965	44	US-10-072-851-6264	Sequence 6264, Ap
41	35.6	7.8	1965	48	US-10-282-122A-20532	Sequence 20532, A
C 42	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 24330, A
C 43	35.4	7.8	3412	1	PCT-US01-08631-24948	Sequence 24948, A
44	34.6	7.6	438	31	US-09-739-449-3383	Sequence 3383, Ap
45	34.6	7.6	438	33	US-09-803-110-3383	Sequence 3383, Ap

ALIGNMENTS

```
RESULT 1
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-139;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-642A-56

  Query Match             88.4%;  Score 403.2;  DB 6;  Length 456;
  Best Local Similarity   92.8%;  Pred. No. 6.1e-122;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||   | | | |  |   | |   |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             88.4%;  Score 403.2;  DB 23;  Length 456;
  Best Local Similarity   92.8%;  Pred. No. 6.1e-122;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||   | | | |  |   | |   |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             78.9%;  Score 360;  DB 23;  Length 456;
  Best Local Similarity   86.8%;  Pred. No. 1.2e-107;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
```

```
Db        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy       241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db       241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 77.5%; Score 353.6; DB 23; Length 456;
Best Local Similarity 86.0%; Pred. No. 1.6e-105;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy        61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
Db        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy       241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db       241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.7%; Pred. No. 5.3e-105;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy        61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
Db        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy       241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db       241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db       421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.7%; Pred. No. 5.3e-105;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db         1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy        61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
Db        61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db       121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db       181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy       241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db       241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db       301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy       361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db       361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 5.3e-105;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||  | | | |   | |     |  ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||        |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 5.3e-105;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||  | | | |   | |     |  ||   ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||      ||  ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||  |    || |  | ||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 5.3e-105;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||   | | | |   | |     |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 6.1e-104;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||   | | | |   | |     |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
                | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             76.5%;  Score 348.8;  DB 23;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 6.1e-104;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||   | | | |   | |     |  ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||  |   |  | ||  |   |   |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

  Query Match             75.8%;  Score 345.6;  DB 23;  Length 456;
  Best Local Similarity   84.9%;  Pred. No. 6.9e-103;
  Matches  387;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              ||||||  |   ||  ||    | | | |  |   | |     |   ||  ||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  ||  |||  |   |   |                 |   | |  ||   ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-642A-54

  Query Match             62.3%;  Score 284.2;  DB 6;  Length 361;
  Best Local Similarity   99.0%;  Pred. No. 1.3e-82;
  Matches  286;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        110 ATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 106

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 166
```

```
Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 226

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 TCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCC 286

Qy        350 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 AATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. Coli
US-09-543-407-3

Query Match 48.8%; Score 222.4; DB 23; Length 456;
Best Local Similarity 68.0%; Pred. No. 3.9e-62;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||   |  ||||   |   | | |   |  ||     |  ||     ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| ||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

RESULT 15
US-09-252-691-172

; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-172

Query Match 48.8%; Score 222.4; DB 17; Length 477;
Best Local Similarity 68.0%; Pred. No. 4e-62;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||| |||||||||   | ||||||||||||||||||||||||| | ||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || |        ||  | |    | |    | |             |          ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || ||||||||  |||||||||| ||||||||   || |||| | |||||||||  ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  || |||||  | ||||||  |||||||||||||   |||  |||||||||||| |||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||| | ||| |||| || |||   || || || ||  | || |
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||| |||||    ||| |||| | ||| |  ||     || | ||||||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||||  |||  | ||||| || || |  || |  || |  ||||  ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||||||||||| || ||  ||||||| | |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477
```

Search completed: March 19, 2004, 03:01:51
Job time : 2652.77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	222.4	48.8	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	180	39.5	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33.4	7.3	3990	5	US-09-830-230A-629	Sequence 629, App
	4	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
c	5	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
c	6	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
c	7	29.4	6.4	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	8	29.4	6.4	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	9	29.4	6.4	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	10	29.4	6.4	1400	7	US-60-545-213-6407	Sequence 6407, Ap
c	11	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	12	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
c	13	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	14	28.4	6.2	2511	1	PCT-US04-05654-2123	Sequence 2123, Ap
c	15	28.4	6.2	20043	6	US-10-021-698A-869	Sequence 869, App
	16	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	17	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
c	18	28.2	6.2	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	19	28.2	6.2	3416	6	US-10-100-683-3397	Sequence 3397, Ap
c	20	28.2	6.2	8206	6	US-10-021-698A-3529	Sequence 3529, Ap
c	21	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
c	22	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
c	23	27.8	6.1	639	6	US-10-767-701-7904	Sequence 7904, Ap
	24	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
c	25	27.8	6.1	2342	6	US-10-451-467A-277	Sequence 277, App
	26	27.8	6.1	2895	6	US-10-767-701-14458	Sequence 14458, A
	27	27.8	6.1	18802	6	US-10-767-471-10574	Sequence 10574, A
c	28	27.8	6.1	32372	6	US-10-767-471-10611	Sequence 10611, A
c	29	27.8	6.1	1126118	6	US-10-767-471-10615	Sequence 10615, A
c	30	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	31	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
	32	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
	33	27.6	6.1	1498	6	US-10-100-683-5108	Sequence 5108, Ap
	34	27.6	6.1	1523	6	US-10-100-683-750	Sequence 750, App
	35	27.6	6.1	1662	6	US-10-100-683-5106	Sequence 5106, Ap
c	36	27.6	6.1	2571	5	US-09-999-183A-4	Sequence 4, Appli
	37	27.6	6.1	3035	6	US-10-786-892-279	Sequence 279, App
c	38	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	39	27.4	6.0	201	6	US-10-767-471-11899	Sequence 11899, A
	40	27.4	6.0	868	6	US-10-767-701-167	Sequence 167, App
c	41	27.4	6.0	318752	6	US-10-767-471-10553	Sequence 10553, A
	42	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
	43	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli
c	44	27	5.9	511	6	US-10-767-701-18763	Sequence 18763, A
c	45	27	5.9	987	6	US-10-100-683-3090	Sequence 3090, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 48.8%; Score 222.4; DB 5; Length 675;
Best Local Similarity 68.0%; Pred. No. 9e-67;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||  | ||||  |   | | | |  |  ||   |  ||    ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 39.5%; Score 180; DB 5; Length 396;
Best Local Similarity 70.6%; Pred. No. 2.8e-52;
Matches 240; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```
Qy        117 CGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCT 176
Db         57 CGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCT 116
Qy        177 GCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGC 236
Db        117 GCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGC 176
Qy        237 CGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAA 296
Db        177 AGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAA 236
Qy        297 TAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGG 356
```

```
Db        237 CAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGG 296
Qy        357 CGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCA 416
Db        297 TGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCA 356
Qy        417 GGTTGGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        357 GGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
; LENGTH: 3990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1139)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1148)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1210)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1244)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1247)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1250)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1251)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

Query Match 7.3%; Score 33.4; DB 5; Length 3990;
Best Local Similarity 47.4%; Pred. No. 0.24;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              |||  ||||||  ||||     |    ||||      ||  ||| |  | | |  |  | |
Db       3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  ||  ||  ||    | |   |  | |||||||||  | |||    |||  ||
Db       3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   ||  |||  |    |    |  |||  |           | |||  |||||  |      |
Db       3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 3612

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
               ||  ||||        |  |  |||  |  ||  |
Db       3613 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 3643


RESULT 4
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
;  APPLICANT: Neubauer, Antonie
;  APPLICANT:  Ziegler, Christina
;  TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
;  FILE REFERENCE: 1/1372
;  CURRENT APPLICATION NUMBER: US/10/624,149A
;  CURRENT FILING DATE:  2003-07-21
;  PRIOR APPLICATION NUMBER: 60/403,282
;  PRIOR FILING DATE: 2002-08-14
;  PRIOR APPLICATION NUMBER: DE 10233064
;  PRIOR FILING DATE: 2002-07-19
;  PRIOR APPLICATION NUMBER: DE 10317008
;  PRIOR FILING DATE: 2003-04-11
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
;   LENGTH: 150223
;   TYPE: DNA
;   ORGANISM: Equine herpesvirus 1
;  PUBLICATION INFORMATION:
;   AUTHORS: Telford,E.A.
;   AUTHORS: Watson, M.S.
;   AUTHORS: McBride, K.
;   AUTHORS: Davison, A.J.
;   TITLE: The DNA sequence of equine herpesvirus-1
;   JOURNAL: Virology
;   VOLUME: 189
;   ISSUE: 1
;   PAGES: 304-316
;   DATE: JUL-1992
;   DATABASE ACCESSION NUMBER: NC 001491, NCBI
;   DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 150223;
  Best Local Similarity   51.4%;  Pred. No. 2.9;
  Matches   75;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              || ||  ||  |  | || ||  | ||| ||| |   |  ||| |||   | ||   |||
Db      95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy        233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
              |       | |||| ||  ||||||     | || |  | | | ||| |    | |||
Db      95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy        293 GAAATAATGCCACCATCGACCAGTGG 318
              || |   |      || ||| | || |
Db      95662 GAGAGCTTATTCACAACGAGCTGTTG 95687


RESULT 5
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
;  APPLICANT: Bayer Pharmaceuticals Corporation
;  APPLICANT:  Eveleigh, Deepa
;  APPLICANT:  Bigwood, Douglas
;  TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
;  FILE REFERENCE: 5152
;  CURRENT APPLICATION NUMBER: US/10/788,792
;  CURRENT FILING DATE:  2004-02-27
;  PRIOR APPLICATION NUMBER: US 60/450,655
;  PRIOR FILING DATE: 2003-02-28
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
;   LENGTH: 1575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-788-792-27

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 1575;
  Best Local Similarity   55.7%;  Pred. No. 1.2;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
              |||   | | | | || ||   | | || |  || | |||     || |||| ||||||||
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
              |  | |   | |  |||  |  |   |   |||  |  ||||||||
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583


RESULT 6
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
;   LENGTH: 24900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(24900)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 24900;
  Best Local Similarity   52.4%;  Pred. No. 11;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||  | ||  ||  | || |  ||   |       |   | || ||  |  | | ||||| |
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |  |||||  |  | | ||||||    | ||||| ||| | | | |      || |    |||
Db       9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929

Qy        327 AAAC 330
              ||||
Db       9928 AATC 9925


RESULT 7
```

US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2134
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-2134

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db 437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy 272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db 377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy 332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
Db 317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy 392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
Db 257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199

RESULT 8
US-60-545-213-2135/c
; Sequence 2135, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2135
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-2135

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db 437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy 272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db 377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy 332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
Db 317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy 392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
Db 257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199

RESULT 9
US-60-545-213-6406/c
; Sequence 6406, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6406
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-6406

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db 437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy 272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db 377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy 332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
Db 317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy 392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
Db 257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199

RESULT 10
US-60-545-213-6407/c
; Sequence 6407, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6407
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-6407

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db 437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

```
Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              |  ||   ||     |      | || ||||    | | |      |         ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              |||  |    | | |||     | |||  |||   |||  |  |||||  |||||       |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |      | | ||  ||    ||  |   |  ||  | |    |  | ||| |
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199


RESULT 11
US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-117

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 249;
  Best Local Similarity   52.5%;  Pred. No. 1.8;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||       || || ||| |    | |  | |   || | | |||
Db        180 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 121

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||     |  |  ||  || || || ||| |  | |||     |||| | | |     || |
Db        120 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 61

Qy        415 CA 416
              ||
Db         60 GA 59


RESULT 12
US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 757
;   TYPE: DNA
;   ORGANISM: E. Coli
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(757)
;   OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 757;
  Best Local Similarity   52.5%;  Pred. No. 3.1;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||       || || ||| |    | |  | |   || | | |||
Db        264 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 323

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||     |  |  ||  || || || ||| |  | |||     |||| | | |     || |
Db        324 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 383

Qy        415 CA 416
              ||
Db        384 GA 385


RESULT 13
US-10-417-375A-143/c
; Sequence 143, Application US/10417375A
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375A
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143
;   LENGTH: 83009
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(83009)
;   OTHER INFORMATION: n = A,T,C or G
US-10-417-375A-143

  Query Match             6.3%;  Score 28.8;  DB 6;  Length 83009;
  Best Local Similarity   48.8%;  Pred. No. 38;
  Matches   78;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        281 AGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTA 340
              |||||  |   ||||   ||   | | || |  ||   |  |   | | |||  ||||
Db      14056 AGAATATTAAAAGAATGCATAATATCCTCAAAAAGAACATTGACTAGTATTCCTTTATTC 13997

Qy        341 CTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCA 400
              || ||    |       | ||||    |  |       || ||| |  |   |  ||| |
Db      13996 CTTTCAAGTACACTTAAGTTAATGCTTCAATGAAATACAAGCAGGCACCTATTACTTCTA 13937

Qy        401 GCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCAC 440
              |  |||||   | |||| | | ||   |  || | |||||
Db      13936 GAATAATGAGACCTCAGATGGATTACAGGCACGATGCCAC 13897


RESULT 14
PCT-US04-05654-2123
; Sequence 2123, Application PC/TUS0405654
; GENERAL INFORMATION:
;  APPLICANT: Sherman, Bradley K
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline  E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
```

```
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Dubell III, Arnold N
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Repetti, Peter
;  APPLICANT:  Century, Karen
;  APPLICANT:  Gutterson, Neal
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Kumimoto, Roderick W
;  APPLICANT:  Pilgrim, Marsha L
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 PCT
;  CURRENT APPLICATION NUMBER: PCT/US04/05654
;  CURRENT FILING DATE:  2004-03-03
;  PRIOR APPLICATION NUMBER: 10/374,780
;  PRIOR FILING DATE: 2003-02-25
;  PRIOR APPLICATION NUMBER: 10/675,852
;  PRIOR FILING DATE: 2003-09-30
;  NUMBER OF SEQ ID NOS: 2950
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2123
;   LENGTH: 2511
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: G1548  Predicted polypeptide sequence is paralogous to G390, G391
PCT-US04-05654-2123

  Query Match             6.2%;  Score 28.4;  DB 1;  Length 2511;
  Best Local Similarity   56.4%;  Pred. No. 10;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         21 AGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGT 80
              ||||   ||||  |  ||||   || | ||||| |  | ||| |  | ||   || |||
Db       1470 AGCAGCAGCAGTTAAAGTAGGGCCTTGTAGTGCCCGAGTTGGAGGATTTGGAGGGCAGGT 1529

Qy         81 TACCCGTGTTGTTACCCATGAAATGGCACATGCA 114
              ||  | |    || | |||   || |  |||| |
Db       1530 TATACTTCCACTTGCTCATACTATTGAGCATGAA 1563


RESULT 15
US-10-021-698A-869/c
; Sequence 869, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 869
;   LENGTH: 20043
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (6103)
;   OTHER INFORMATION: a, t, c or g
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (20008)
;   OTHER INFORMATION: a, t, c or g
US-10-021-698A-869

  Query Match             6.2%;  Score 28.4;  DB 6;  Length 20043;
  Best Local Similarity   56.4%;  Pred. No. 27;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          9 TTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCTA 68
              | ||| | ||  |   | ||| || || || |  || ||| || | | | |||| | ||
Db      10361 TCTAATACTGTAATACTGCGCTGCTATTGTTGGCTCCGGCTGTCCACAGCCTGGGGGCTG 10302

Qy         69 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAA 102
                ||  ||   | | || |||    | | |  |||
Db      10301 GGAAGAGAGTGCTGCCTGTGGAAATGCTCGGGAA 10268


Search completed: March 19, 2004, 03:10:14
Job time : 57.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	223	48.9	680	12	BJ618688	BJ618688 BJ618688
c	2	37.4	8.2	484	9	AV430994	AV430994 AV430994
c	3	37.2	8.2	772	29	CG862779	CG862779 ZMMBBc027
c	4	35.8	7.9	484	14	CF089570	CF089570 QHM23E04.
	5	35.2	7.7	813	28	BZ204853	BZ204853 CH230-390
	6	34.6	7.6	702	12	BG441624	BG441624 GA__Ea001
	7	34.6	7.6	937	12	BG444255	BG444255 GA__Ea002
	8	34.2	7.5	296	9	AV058630	AV058630 AV058630
	9	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
c	10	34	7.5	811	28	BZ469000	BZ469000 BOOAO58TF
c	11	33.8	7.4	480	14	CF372418	CF372418 CSECS052B
c	12	33.8	7.4	481	14	CF414539	CF414539 CSECS071B
	13	33.8	7.4	509	14	CF605114	CF605114 RADIC01_0
	14	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	15	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
c	16	33.6	7.4	712	14	CF437982	CF437982 EST674327
	17	33.4	7.3	679	29	CC961526	CC961526 BOIGS10TF
c	18	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
c	19	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF
c	20	33.4	7.3	709	28	BH690914	BH690914 BOHXR64TF
c	21	33.4	7.3	785	28	BH578793	BH578793 BOGLY20TR
c	22	33.4	7.3	815	28	BH461716	BH461716 BOHIV03TR
c	23	33.4	7.3	827	28	BH438448	BH438448 BOGVQ28TR
c	24	33.4	7.3	830	28	BH715216	BH715216 BOMIM42TR
c	25	33.4	7.3	835	28	BH575194	BH575194 BOHJO27TF
c	26	33.4	7.3	853	28	BH578402	BH578402 BOHNV61TF
c	27	33.4	7.3	1026	28	BZ458784	BZ458784 BONHB14TF
	28	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
	29	33.2	7.3	625	10	BF501499	BF501499 AT16857.5
	30	33.2	7.3	788	9	AU139321	AU139321 AU139321
	31	33	7.2	456	12	BJ331339	BJ331339 BJ331339
	32	33	7.2	524	12	BJ367372	BJ367372 BJ367372
	33	33	7.2	547	12	BJ333958	BJ333958 BJ333958
	34	33	7.2	553	12	BJ336903	BJ336903 BJ336903
	35	33	7.2	569	12	BJ334047	BJ334047 BJ334047
	36	33	7.2	574	12	BJ330328	BJ330328 BJ330328
	37	33	7.2	607	12	BJ365766	BJ365766 BJ365766
	38	33	7.2	625	12	BJ396922	BJ396922 BJ396922
	39	33	7.2	629	12	BJ335378	BJ335378 BJ335378
	40	33	7.2	635	12	BJ397391	BJ397391 BJ397391
	41	33	7.2	645	12	BJ392753	BJ392753 BJ392753
	42	33	7.2	670	12	BJ336637	BJ336637 BJ336637
c	43	32.8	7.2	518	10	BF755518	BF755518 MR2-CT059
	44	32.8	7.2	572	12	BI626794	BI626794 RH67848.5
	45	32.8	7.2	574	10	BF854273	BF854273 MR2-EN009

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.

```
FEATURES     http://xenopus.nibb.ac.jp.
     source          Location/Qualifiers
                     1. .680
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match                    48.9%;  Score 223;  DB 12;  Length 680;
  Best Local Similarity   68.0%;  Pred. No. 3.1e-58;
  Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138

RESULT 2
AV430994/c
LOCUS
DEFINITION  AV430994
            AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PL027e10_r 5', mRNA sequence.
ACCESSION   AV430994
VERSION     AV430994.1  GI:8586219
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
  PUBMED    10907854
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/
FEATURES
     source          Location/Qualifiers
                     1. .484
                     /organism="Porphyra yezoensis"
                     /mol_type="mRNA"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PL027e10_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2
                     XhoI"
ORIGIN

  Query Match                    8.2%;  Score 37.4;  DB 9;  Length 484;
  Best Local Similarity   57.1%;  Pred. No. 2.4;
  Matches   68;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323

RESULT 3
CG862779/c
LOCUS
DEFINITION  CG862779
            ZMMBBc0273B12f ZMMBBc (EcoRI) Zea mays subsp. mays genomic survey sequence.
ACCESSION   CG862779
VERSION     CG862779.1  GI:38436806
KEYWORDS    GSS.
SOURCE      Zea mays subsp. mays (maize)
  ORGANISM  Zea mays subsp. mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003c)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
FEATURES    High quality sequence start: 450.
     source          Location/Qualifiers
                     1. .772
                     /organism="Zea mays subsp. mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /sub_species="mays"
                     /db_xref="taxon:4578"
                     /clone="ZMMBBc0273B12"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBc (EcoRI)"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"
ORIGIN

  Query Match                    8.2%;  Score 37.2;  DB 29,  Length 772,
  Best Local Similarity   51.9%;  Pred  No  3 6,
  Matches   84;  Conservative    0,  Mismatches   78,  Indels    0,  Gaps    0,
```

```
Qy          45 TGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAAT 104
               || |  |   || ||| ||      ||     || |   |||   ||| || ||| |
Db         202 TGTCCATCAGCTCGCTCGCCCTGGTGGGGGTCTTGCGTCCCAACTTGCTAACCAGGTCGC 143

Qy         105 GGCACATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
               ||||  ||  ||||||  |||| |  || |||||||| | || ||| | |          ||
Db         142 GGCAGGTGGTTCCGGCAAGGACACGCCGATGAGCATCTGTCGGTATGTTGGGCAGCTCGG 83

Qy         165 TGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGAC 206
               || |||   | || | |   |     |||   | | |  |||
Db          82 TGAGCTGCATTTGAACATTCAGCAATGTAGTCCAGGAGAGAC 41


RESULT 4
CF089570/c
LOCUS       CF089570               484 bp    mRNA    linear   EST 22-JUL-2003
DEFINITION  QHM23E04.yg.ab1 QH_M sunflower H.argophyllus Helianthus argophyllus
            cDNA clone QHM23E04, mRNA sequence.
ACCESSION   CF089570
VERSION     CF089570.1  GI:33128637
KEYWORDS    EST.
SOURCE      Helianthus argophyllus
  ORGANISM  Helianthus argophyllus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QH_CA_Contig2049, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QHM23  row: E  column: 04.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Helianthus argophyllus"
                     /mol_type="mRNA"
                     /db_xref="taxon:73275"
                     /clone="QHM23E04"
                     /lab_host="E.coli"
                     /clone_lib="QH_M sunflower H.argophyllus"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from three different sources (seedling, root and leaf) of
                     RNA from a single genotype. cDNAs were pooled and
                     directionally cloned into a custom medium-copy vector.
                     Details of library construction can be obtained at
                     http://cgpdb.ucdavis.edu/"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 14;  Length 484;
  Best Local Similarity   46.0%;  Pred. No. 7.4;
  Matches  121;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy          58 GCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCC 117
               |||||  ||  |||||  |  | |||| ||  || ||     ||| |    ||| |  |
Db         323 GCTGGGTTCATTGATCCTCCTGATACCGGTCCTGCTAAGATGGAAGTAATGCATACGGCT 264

Qy         118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
               |  ||     ||     |   | ||     || |      |  ||    |||
Db         263 GATCCTACTGCCATACCGGCTAGCTCTCCTACGGCTTTAGAATCTGTAGCTACGGCTGAA 204

Qy         178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
               |     |   | | |     | || |  ||| || || | |   | |||||   | |
Db         203 GTGACAAACATCATTGAGAAAGTTACAATGATTTCCATGATCAGAGCTTGTAACTCTGTC 144

Qy         238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
               | ||||  || || |  ||    | ||||| |   |     |   ||  |  ||||
Db         143 CCAGAAGGCGTGGTTGTGCCTAGATTTGTTATTGGATCTAACATTATCCGTAGCGCAAAT 84

Qy         298 AATGCCACCATCGACCAGTGGAA 320
               | ||||  || ||||    |||
Db          83 GAAGCCAAGATGGACCCTGTGAA 61


RESULT 5
BZ204853
LOCUS       BZ204853               813 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-390D21, genomic survey sequence.
ACCESSION   BZ204853
VERSION     BZ204853.1  GI:23862905
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-390D21.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 390  row: D  column: 21
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-390D21"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             7.7%;  Score 35.2;  DB 28;  Length 813;
  Best Local Similarity   49.0%;  Pred. No. 16;
  Matches   94;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy         178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
               |||  |   | | | | || |   |    ||    ||  | ||  |  |   ||  | ||
Db         548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy         238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
```

```
Db        608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357

Db        668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAAGGACAC 727

Qy        358 GGTAATAACGCC 369

Db        728 TGTGGTTAGGAC 739
```

RESULT 6
BG441624
LOCUS BG441624 702 bp mRNA linear EST 15-MAR-2001
DEFINITION GA__Ea0014A14f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA__Ea0014A14f, mRNA sequence.
ACCESSION BG441624
VERSION BG441624.1 GI:13351276
KEYWORDS EST.
SOURCE Gossypium arboreum
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE 1 (bases 1 to 702)
AUTHORS Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 696.
FEATURES Location/Qualifiers
source 1. .702
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA__Ea0014A14f"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 702;
Best Local Similarity 46.5%; Pred. No. 22;
Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

```
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        185 AACAATGACGGGAATGCGAATGGGAATGTTGGCAGCACAGACAATCAGAGCAACACCAAT 244

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        245 GACAATGTAAATGCAAATGAAGATGTTACTCAAAAGGAAAACAACGATTCAAGTTCAAAT 304

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        305 GAAAATGCTGCACAGAGCAACACCAACAACAATGAAAACAATAACAATGAAAATGCGGGT 364

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        365 CAGGACGGGAAGAATGCCACTCAGAGTAACACTGACAACAATGAAAATGTGGGTCAGAAT 424

Qy        421 G 421

Db        425 G 425
```

RESULT 7
BG444255
LOCUS BG444255 937 bp mRNA linear EST 15-MAR-2001
DEFINITION GA__Ea0023M14f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA__Ea0023M14f, mRNA sequence.
ACCESSION BG444255
VERSION BG444255.1 GI:13353907
KEYWORDS EST.
SOURCE Gossypium arboreum
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE 1 (bases 1 to 937)
AUTHORS Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 804.
FEATURES Location/Qualifiers
source 1. .937
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA__Ea0023M14f"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 937;
Best Local Similarity 46.5%; Pred. No. 26;
Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

```
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        185 AACAATGACGGGAATGCGAATGGGAATGTTGGCAGCACAGACAATCAGAGCAACACCAAT 244

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        245 GACAATGTAAATGCAAATGAAGATGTTACTCAAAAGGAAAACAACGATTCAAGTTCAAAT 304

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        305 GAAAATGCTGCACAGAGCAACACCAACAACAATGAAAACAATAACAATGAAAATGCGGGT 364

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        365 CAGGACGGGAAGAATGCCACTCAGAGTAACACTGACAACAATGAAAATGTGGGTCAGAAT 424

Qy        421 G 421

Db        425 G 425
```

```
RESULT 8
AV058630
LOCUS       AV058630                296 bp    mRNA    linear   EST 23-JUN-1999
DEFINITION  AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA
            clone 1810054H02, mRNA sequence.
ACCESSION   AV058630
VERSION     AV058630.1  GI:5158377
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 296)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .296
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1810054H02"
                     /sex="male"
                     /tissue_type="pancreas"
                     /dev_stage="adult"
                     /clone_lib="Mus musculus pancreas C57BL/6J adult"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 9;  Length 296;
  Best Local Similarity   53.3%;  Pred. No. 17;
  Matches   72;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCG 310
              |||  ||  | |||  | |  |  | ||  ||  ||| |  |||||  | ||  | |  |
Db         13 GTGGAGACCACAGTTTTTTCTAGGTCACCAAGGCTGGGTAAAGAAAAGAAGCATCAACTG 72

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              |||||| | | | |||||    ||    | ||    || | | |  || | ||| |  |
Db         73 TCCAGTGAAGCACAAAAAATGAAGAAGAAAATGATTACCCAAAAGTTGGAATTAAGGTAG 132

Qy        371 CGCTGGTTAATCAGA 385
              | ||   ||  |||
Db        133 AGATGCCAAAATAGA 147


RESULT 9
BJ334624
LOCUS       BJ334624                523 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda47o06 5', mRNA sequence.
ACCESSION   BJ334624
VERSION     BJ334624.1  GI:19164754
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda47o06"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 523;
  Best Local Similarity   57.1%;  Pred. No. 25;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        398 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGANGACAATAAT 457

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACT 342
              |||  || || | || |    ||   ||||||     |||||| |
Db        458 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAAT 502


RESULT 10
BZ469000/c
LOCUS       BZ469000                811 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BOOAO58TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOAO58,
            genomic survey sequence.
ACCESSION   BZ469000
VERSION     BZ469000.1  GI:26764546
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOOAO58TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .811
```

/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOOAO58"
/clone_lib="BO_1.6_2_KB_tot"
/note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match 7.5%; Score 34; DB 28; Length 811;
Best Local Similarity 48.9%; Pred. No. 37;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```
Qy        176 TGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCG 235
              ||  | || |     ||||| |||       |||||      |  | |||||  ||  ||
Db        610 TGTTAGGCAAGCATTGTAAAACTGTGTTTCCCATTTCATCTACTGTTTATGACAAATGCT 551

Qy        236 CCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
                ||| ||   ||   ||  | |  | | ||  |    || ||  | || |||| | |
Db        550 TTGATCTAATTCACTCTGATGTTTGGACTGCTCCTTGCTTGTCTAGGGATAGTTTTAAAT 491

Qy        296 ATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 355
              |   || |||  ||    |    ||  | ||| || ||  || ||  |    ||| |
Db        490 ACTTTGTCACATTCATAAATGAAAAATCGAAATACACCTGGATAACACTTATTCAAACCA 431

Qy        356 GCGGTA 361
                | ||
Db        430 AAGATA 425
```

RESULT 11
CF372418/c
LOCUS CF372418 480 bp mRNA linear EST 27-AUG-2003
DEFINITION CSECS052B08_FLOn0012 CabSau Normalised Flower Stage 12 (FLOn0012) Vitis vinifera cDNA clone CSECS052B08 3', mRNA sequence.
ACCESSION CF372418
VERSION CF372418.1 GI:34319664
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 480)
AUTHORS Iocco,P., Hua,C., Davies,C. and Thomas,M.R.
TITLE Expressed sequence tags from the grapevine cultivar Cabernet Sauvignon
JOURNAL Unpublished (2003)
COMMENT Contact: Mark R. Thomas
CSIRO Plant Industry
CSIRO
PO Box 350, Glen Osmond, SA, 5064, Australia
Tel: 61 8 83038600
Fax: 61 8 83038601
Email: Mark.R.Thomas@csiro.au
Seq primer: CCCAGTCACGACGTTGTAAAACG (M13 Forward)
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .480
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSECS052B08"
/sex="Hermaphrodite"
/dev_stage="12 - modified E-L system"
/clone_lib="CabSau Normalised Flower Stage 12 (FLOn0012)"
/note="Organ: Inflorescence including flowers; Vector: pZL; Normalised cDNA library from immature inflorescences at stage 12 of the modified E-L system. Tissue collected from field grown plants. A description of the modified E-L system can be found in the paper by B. G. Coombe 'Adoption of a system for identifying grapevine growth stages' (1995) Aust. J. Grape and Wine Res. 1: 104-110."

ORIGIN

Query Match 7.4%; Score 33.8; DB 14; Length 480;
Best Local Similarity 49.7%; Pred. No. 31;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy         94 ACCCATGAAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              | | ||||  ||| || ||   |   |||   | ||| |      |      |  || |
Db        222 ATCAATGAGGTGGAACCTGACCCTTCCCCTCCCCCAAAGCAACTGTTGCGAGGGGCCGCT 163

Qy        154 TCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACC 213
              |||||  |||||||  ||||      |||||  | |   |   || |||    |||| ||
Db        162 GCCGCTGCCGCTGCGGCTGCTGCTGCTAGCGAAGACACGAGGGCTAAAAGCCACATTTCC 103

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||||  | |   |    ||   |       ||| || |  |||| ||  | |
Db        102 CAGAAAGAATCAAGGCTTGGAAGCAGCAGTAGCCGGGATCAGGATGATTCTCC 50
```

RESULT 12
CF414539/c
LOCUS CF414539 481 bp mRNA linear EST 02-SEP-2003
DEFINITION CSECS071B09_POSn0036 CabSau Normalised Berry Postveraison Stage 36 (POSn0036) Vitis vinifera cDNA clone CSECS071B09 3', mRNA sequence.
ACCESSION CF414539
VERSION CF414539.1 GI:34415785
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 481)
AUTHORS Iocco,P., Hua,C., Davies,C. and Thomas,M.R.
TITLE Expressed sequence tags from the grapevine cultivar Cabernet Sauvignon
JOURNAL Unpublished (2003)
COMMENT Contact: Mark R. Thomas
CSIRO Plant Industry
CSIRO
PO Box 350, Glen Osmond, SA, 5064, Australia
Tel: 61 8 83038600
Fax: 61 8 83038601
Email: Mark.R.Thomas@csiro.au
Seq primer: CCCAGTCACGACGTTGTAAAACG (M13 Forward)
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .481
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSECS071B09"
/sex="Hermaphrodite"
/dev_stage="36 - modified E-L system"
/clone_lib="CabSau Normalised Berry Postveraison Stage 36 (POSn0036)"
/note="Organ: Fruit with seeds removed; Vector: pZL; A cDNA library from postveraison fruit at stage 36 of the modified E-L system. Tissue collected from field grown plants. A description of the modified E-L system can be found in the paper by B. G. Coombe 'Adoption of a system for identifying grapevine growth stages' (1995) Aust. J. Grape and Wine Res. 1: 104-110."

ORIGIN

Query Match 7.4%; Score 33.8; DB 14; Length 481;
Best Local Similarity 49.7%; Pred. No. 31;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy         94 ACCCATGAAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
```

```
                 | | ||||  ||| || ||   |  |||  | ||| |       |      |  || |
Db        223 ATCAATGAGGTGGAACCTGACCCTTCCCCTCCCCCAAAGCAACTGTTGCGAGGGGCCGCT 164

Qy        154 TCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACC 213
                |||||  |||||||  ||||      |||||  | |   |   || |||    |||| ||
Db        163 GCCGCTGCCGCTGCGGCTGCTGCTGCTAGCGAAGACACGAGGGCTAAAAGCCACATTTCC 104

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
                ||||  |  |   |    ||   |        ||| || | | |||| ||  | |
Db        103 CAGAAAGAATCAAGGCTTGGAAGCAGCAGTAGCCGGGATCAGGATGATTCTCC 51


RESULT 13
CF605114
LOCUS       CF605114                509 bp    mRNA    linear   EST 30-SEP-2003
DEFINITION  RADIC01_000700 Grape Root pSPORT1 Library Vitis vinifera cDNA 5',
            mRNA sequence.
ACCESSION   CF605114
VERSION     CF605114.1  GI:37185761
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Moser,C., Segala,C., Fontana,P., Salakhudtinov,I., Gatto,P.,
            Pindo,M., Zyprian,E., Toepfer,R., Grando,M.S. and Velasco,R.
  TITLE     Expressed sequence tags from different organs of Vitis vinifera
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Moser C
            Laboratorio di Genetica Molecolare
            Istituto Agrario di San Michele all'Adige (IASMA)
            via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia
            Tel: 0039-0461-615314
            Fax: 0039-0461-650956
            Email: claudio.moser@ismaa.it
            The sequencing work has been funded by the 'Fondazione Cassa di
            Risparmio di Trento e Rovereto'
            High quality sequence stop: 509.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Pinot noir"
                     /db_xref="taxon:29760"
                     /sex="Hermaphrodite"
                     /dev_stage="young root (first year)"
                     /lab_host="DH10B"
                     /clone_lib="Grape Root pSPORT1 Library"
                     /note="Organ: root; Vector: pSPORT1; Site_1: NotI; Site_2:
                     SalI"
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 509;
  Best Local Similarity   49.7%;  Pred. No. 32;
  Matches   86;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         94 ACCCATGAAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              | | ||||  ||| || ||   |  |||  | ||| |       |      |  || |
Db        295 ATCAATGAGGTGGAACCTGACCCTTCCCCTCCCCCAAAGCAACTGTTGCGAGGGGCCGCT 354

Qy        154 TCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACC 213
                |||||  |||||||  ||||      |||||  | |   |   || |||    |||| ||
Db        355 GCCGCTGCCGCTGCGGCTGCTGCTGCTAGCGAAGACACGAGGGCTAAAAGCCACATTTCC 414

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
                ||||  |  |   |    ||   |        ||| || | | |||| ||  | |
Db        415 CAGAAAGAATCAAGGCTTGGAAGCAGCAGTAGCCGGGATCAGGATGATTCTCC 467


RESULT 14
BJ335653
LOCUS       BJ335653                594 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ335653 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda51h06 5', mRNA sequence.
ACCESSION   BJ335653
VERSION     BJ335653.1  GI:19165783
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda51h06"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.4%;  Score 33.6;  DB 12;  Length 594;
  Best Local Similarity   57.1%;  Pred. No. 41;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   |||||  | | | |   ||||    ||  |||
Db        223 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 282

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACT 342
              |||  || || | || |    ||   ||||||     |||||| |
Db        283 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAAT 327


RESULT 15
BJ336964
LOCUS       BJ336964                653 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ336964 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda55e20 5', mRNA sequence.
ACCESSION   BJ336964
VERSION     BJ336964.1  GI:19167094
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .653
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
```

```
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda55e20"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match              7.4%;  Score 33.6;  DB 12;  Length 653;
  Best Local Similarity   53.9%;  Pred. No. 43;
  Matches   69;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |     ||| |||||   ||||| | | | |   ||||    ||  |||
Db        187 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 246

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              |||  || || | || |     ||   ||||||      |||||| |    |   | | |
Db        247 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGAAGAA 306

Qy        358 GGTAATAA 365
                |||  ||
Db        307 AATAAAAA 314


Search completed: March 18, 2004, 20:11:28
Job time : 2034.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	406.4	89.1	456	6	I44909	I44909 Sequence 58
2	406.4	89.1	2067	1	SEU43280	U43280 Salmonella
3	400	87.7	5103	1	STAJ2301	AJ002301 Salmonell
4	400	87.7	22411	1	AE008749	AE008749 Salmonell
5	398.4	87.4	254050	1	AL627269	AL627269 Salmonell
c 6	398.4	87.4	301983	1	AE016840	AE016840 Salmonell
7	392.6	86.1	1048	1	STAGFBA	AJ000514 Salmonell
8	282.2	61.9	361	6	I44908	I44908 Sequence 56
9	280	61.4	2889	1	CSP515700	AJ515700 Citrobact
10	245.2	53.8	2920	1	CFR515701	AJ515701 Citrobact
11	232	50.9	4680	1	ECCSGABDG	X90754 E.coli csgG
12	232	50.9	10346	1	AE000205	AE000205 Escherich
13	232	50.9	15047	1	D90741	D90741 Escherichia
14	230.4	50.5	456	6	AX814811	AX814811 Sequence
15	230.4	50.5	648	1	ECOCSGAA	L04979 Escherichia
16	220.6	48.4	306358	1	AE016759	AE016759 Escherich
17	219	48.0	1711	1	AF275733	AF275733 Escherich
18	219	48.0	10190	1	AE005315	AE005315 Escherich
19	219	48.0	327773	1	AP002554	AP002554 Escherich
20	192	42.1	2883	1	ESA515702	AJ515702 Enterobac
21	179.4	39.3	230	1	SEU53207	U53207 Salmonella
22	163.2	35.8	10370	1	AE015131	AE015131 Shigella
23	163.2	35.8	292504	1	AE016981	AE016981 Shigella
c 24	140.2	30.7	437	1	AF237726	AF237726 Shigella
25	93.2	20.4	19201	1	D90742	D90742 Escherichia
26	64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27	48.2	10.6	78	6	AX814809	AX814809 Sequence
28	39.4	8.6	10999	1	AE002536	AE002536 Neisseria
29	39.4	8.6	349980	6	AX044034	AX044034 Sequence
c 30	37.8	8.3	349061	1	NMA2Z2491	AL162753 Neisseria
c 31	37	8.1	9955	1	AE002504	AE002504 Neisseria
c 32	37	8.1	349980	6	AX044033	AX044033 Sequence
c 33	36.8	8.1	126062	8	AP006135	AP006135 Lotus cor
34	36.6	8.0	110000	2	AC142602_0	AC142602 Macaca mu
35	36.6	8.0	305730	2	AC114347	AC114347 Rattus no
36	36.6	8.0	311823	2	AC095362	AC095362 Rattus no
37	36.4	8.0	10798	1	AE004604	AE004604 Pseudomon
38	36.4	8.0	110000	2	AC110929_4	Continuation (5 of
39	36.4	8.0	181996	2	AC108614	AC108614 Rattus no
c 40	36.4	8.0	212174	2	AC105896	AC105896 Rattus no
c 41	36.4	8.0	305724	2	AC128763	AC128763 Rattus no
42	36.2	7.9	1797	3	LEIMSP52A	L19563 Leishmania
43	36.2	7.9	1860	3	LIGP63GEN	Z83677 L.infantum
44	36.2	7.9	2104	3	LDO495002	AJ495002 Leishmani
45	36.2	7.9	2106	3	LDO495006	AJ495006 Leishmani

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           89.1%;  Score 406.4;  DB 6;  Length 456;
  Best Local Similarity  93.2%;  Pred. No. 1.9e-103;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||     |   ||  |    | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match           89.1%;  Score 406.4;  DB 1;  Length 2067;
  Best Local Similarity  93.2%;  Pred. No. 2.1e-103;
```

```
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301                5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYPSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match           87.7%;  Score 400;  DB 1;  Length 5103;
  Best Local Similarity 92.3%;  Pred. No. 1.4e-101;
  Matches  421;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |     |  ||  |    | | ||  | |    || | ||| ||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421

RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
RBS                     complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
gene                    complement(2580. .4088)
                        /gene="STM1128"
CDS                     complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
RBS                     complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
gene                    complement(4413. .5099)
                        /gene="STM1129"
CDS                     complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
RBS                     complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
gene                    5599. .6759
                        /gene="STM1130"
CDS                     5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
gene                    6793. .7497
                        /gene="STM1131"
RBS                     6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           87.7%;  Score 400;  DB 1;  Length 22411;
  Best Local Similarity  92.3%;  Pred. No. 1.6e-101;
  Matches  421;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |     |   ||  |    | | ||  | |    || | ||| ||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
  PUBMED    11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAININSSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSPEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSPSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             87.4%;  Score 398.4;  DB 1;  Length 254050;
  Best Local Similarity   92.1%;  Pred. No. 5.2e-101;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |      |    ||   |     | | ||  | |    || | ||| ||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             87.4%;  Score 398.4;  DB 1;  Length 301983;
Best Local Similarity   92.1%;  Pred. No. 5.3e-101;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |     |   ||   |     | | ||  | |    || | ||| ||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match           86.1%;  Score 392.6;  DB 1;  Length 1048;
  Best Local Similarity  91.4%;  Pred. No. 1.6e-99;
  Matches  416;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |       |    ||   |     | | ||  | |    || | ||| ||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           61.9%;  Score 282.2;  DB 6;  Length 361;
  Best Local Similarity  90.1%;  Pred. No. 1.9e-68;
  Matches  302;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
              ||||       |    ||   |     | | ||  | |    || | ||| |||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700              2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAPAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           61.4%;  Score 280;  DB 1;  Length 2889;
  Best Local Similarity  77.4%;  Pred. No. 9e-68;
  Matches  353;  Conservative    0;  Mismatches  100;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |     ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |     | |    ||   |     |    | || |    || | ||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||| || |||||||||| |||||||||   ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || ||||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || |||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
CFR515701
LOCUS       CFR515701              2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
Query Match             53.8%;  Score 245.2;  DB 1;  Length 2920;
Best Local Similarity   72.8%;  Pred. No. 5.8e-58;
Matches  332;  Conservative    0;  Mismatches  118;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| ||||||||||||       |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              || |||| |   | |    |   ||    ||   | | |    ||| |||| ||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||     ||||| || ||  |||   ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||| ||| | ||||| |||    ||||  |||| || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| |||| |||||||| || |||||  | |||||||  ||    ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||| ||||||||||||| || ||||| ||||  ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || |||||||| | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

```
RESULT 11
ECCSGABDG
LOCUS       ECCSGABDG                 4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
     gene            /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
     CDS             /gene="csgG"
                     complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
     CDS             /gene="csgF"
                     complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
     CDS             /gene="csgE"
                     complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
     CDS             /gene="csgD"
                     complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
misc_feature         complement(2626)
```

```
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           50.9%;  Score 232;  DB 1;  Length 4680;
  Best Local Similarity   69.3%;  Pred. No. 3.1e-54;
  Matches  316;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  | | |    |||| ||| ||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| | |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 12
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
```

```
          and products; all new functional assignments courtesy of Monica
          Riley; added promoters, protein binding sites, and repeated
          sequences described in reference 1. The unique numeric identifiers
          beginning with a lowercase 'b' assigned to each gene (protein- or
          RNA-encoding) are now designated as gene synonyms instead of
          labels. This should allow them to be searched for in Entrez as gene
          names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

Query Match          50.9%;  Score 232;  DB 1;  Length 10346;
Best Local Similarity  69.3%;  Pred. No. 3.3e-54;
Matches  316;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||||  ||||| || || || || ||    ||.| ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

RESULT 13
D90741
LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
FEATURES Location/Qualifiers
source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER

```
                ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS             2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
CDS             3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
gene            complement(4117. .4950)
                /gene="csgG"
CDS             complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
gene            complement(4977. .5393)
                /gene="csgF"
CDS             complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene            complement(5418. .5807)
                /gene="csgE"
CDS             complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
gene            complement(5812. .6462)
                /gene="csgD"
CDS             complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene            7217. .7672
                /gene="csgB"
CDS             7217. .7672
```

Query Match 50.9%; Score 232; DB 1; Length 15047;
Best Local Similarity 69.3%; Pred. No. 3.4e-54;
Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    || ||      | |  |  | |    |||| ||| ||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||.|| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168
```

```
RESULT 14
AX814811
LOCUS       AX814811                456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
```

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           50.5%;  Score 230.4;  DB 6;  Length 456;
  Best Local Similarity  69.1%;  Pred. No. 7.6e-54;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                 648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match           50.5%;  Score 230.4;  DB 1;  Length 648;
  Best Local Similarity  69.1%;  Pred. No. 7.8e-54;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 18, 2004, 15:02:51
Job time : 2312.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64625	Aac64625 AgfA::PT3
	2	406.4	89.1	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	406.4	89.1	456	2	AAT74142	Aat74142 Salmonell
	4	406.4	89.1	456	3	AAC64617	Aac64617 Salmonell
	5	363.2	79.6	456	3	AAC64626	Aac64626 AgfA::PT3
	6	355.2	77.9	456	3	AAC64628	Aac64628 AgfA::PT3
	7	355.2	77.9	456	3	AAC64622	Aac64622 AgfA::PT3
	8	355.2	77.9	456	3	AAC64629	Aac64629 AgfA::PT3
	9	355.2	77.9	456	3	AAC64623	Aac64623 AgfA::PT3
	10	353.6	77.5	456	3	AAC64624	Aac64624 AgfA::PT3
	11	352	77.2	456	3	AAC64630	Aac64630 AgfA::PT3
	12	352	77.2	456	3	AAC64627	Aac64627 AgfA::PT3
	13	348.8	76.5	456	3	AAC64631	Aac64631 AgfA::PT3
	14	282.2	61.9	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	282.2	61.9	361	2	AAT74141	Aat74141 Salmonell
	16	230.4	50.5	456	3	AAC64619	Aac64619 Escherich
	17	230.4	50.5	456	9	ACF36153	Acf36153 E. coli C
	18	218	47.8	646	2	AAQ62647	Aaq62647 Fibronect
	19	156	34.2	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
c	21	50	11.0	78	3	AAC64609	Aac64609 AgfA (SEF
	22	50	11.0	78	3	AAC64610	Aac64610 AgfA (SEF
	23	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	24	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	25	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	26	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
	27	41.2	9.0	78	3	AAC64606	Aac64606 SefA (SEF
	28	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
c	29	40.6	8.9	78	3	AAC64605	Aac64605 SefA (SEF
	30	39.4	8.6	47475	3	AAA81465	Aaa81465 N. mening
	31	39.4	8.6	110000	3	AAA81489_5	Continuation (6 of
	32	39.4	8.6	349980	3	AAF21612	Aaf21612 Neisseria
	33	39.2	8.6	1083	5	AAS76745	Aas76745 DNA encod
c	34	37	8.1	14652	3	AAA81482	Aaa81482 N. mening
c	35	37	8.1	110000	3	AAA81489_1	Continuation (2 of
c	36	37	8.1	349980	3	AAF21611	Aaf21611 Neisseria
c	37	36.4	8.0	2610	7	ACA42270	Aca42270 Prokaryot
c	38	36	7.9	2667	7	ACA45315	Aca45315 Prokaryot
c	39	35.8	7.9	954	7	ACA25910	Aca25910 Prokaryot
	40	35.6	7.8	2000	7	ADA71938	Ada71938 Rice gene
	41	35.4	7.8	3300	6	ABS67377	Abs67377 Neisseria
c	42	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	43	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
c	44	34.8	7.6	2000	7	ADA71938	Ada71938 Rice gene
	45	34.6	7.6	456	3	AAC64620	Aac64620 Escherich

ALIGNMENTS

RESULT 1
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.3e-135;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 2; Length 456;
Best Local Similarity 93.2%; Pred. No. 9.9e-120;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||  |   ||  |   | | ||  | |   || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match            89.1%;  Score 406.4;  DB 2;  Length 456;
  Best Local Similarity  93.2%;  Pred. No. 9.9e-120;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||     |   ||   |   | || |  |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

```
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             89.1%;  Score 406.4;  DB 3;  Length 456;
  Best Local Similarity   93.2%;  Pred. No. 9.9e-120;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match             79.6%;  Score 363.2;  DB 3;  Length 456;
  Best Local Similarity   87.3%;  Pred. No. 6.9e-106;
  Matches  398;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||| |||   |    ||||||     | |    | |    ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match           77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity  86.2%;  Pred. No. 2.5e-103;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||       ||    ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.5e-103;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |   ||  |    | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||| |         | || || ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.5e-103;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |   ||  |    | | ||  | |    || | |    ||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |         ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 9
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.5e-103;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||          |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 10
AAC64624
ID   AAC64624 standard; DNA; 456 BP.
XX
AC   AAC64624;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 8.2e-103;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |  ||  ||   | | | |  |  | |    |  ||   ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |   ||  |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match            77.2%;  Score 352;  DB 3;  Length 456;
  Best Local Similarity  85.7%;  Pred. No. 2.7e-102;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | |||   |      | | ||  |    |      ||| ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            77.2%;  Score 352;  DB 3;  Length 456;
  Best Local Similarity  85.7%;  Pred. No. 2.7e-102;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||  |  |       | || |    || |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity   85.3%;  Pred. No. 2.8e-101;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||||| |  |  ||  |  | | ||  | |   || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |   |                |   | |  ||   ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match            61.9%;  Score 282.2;  DB 2;  Length 361;
  Best Local Similarity  90.1%;  Pred. No. 5.6e-80;
  Matches  302;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
              ||||       |    ||   |      | | ||  | |    || | ||| |||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match            61.9%;  Score 282.2;  DB 2;  Length 361;
  Best Local Similarity  90.1%;  Pred. No. 5.6e-80;
  Matches  302;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
              ||||       |    ||   |      | | ||  | |    || | ||| |||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 18, 2004, 09:11:59
```

Job time : 290.405 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

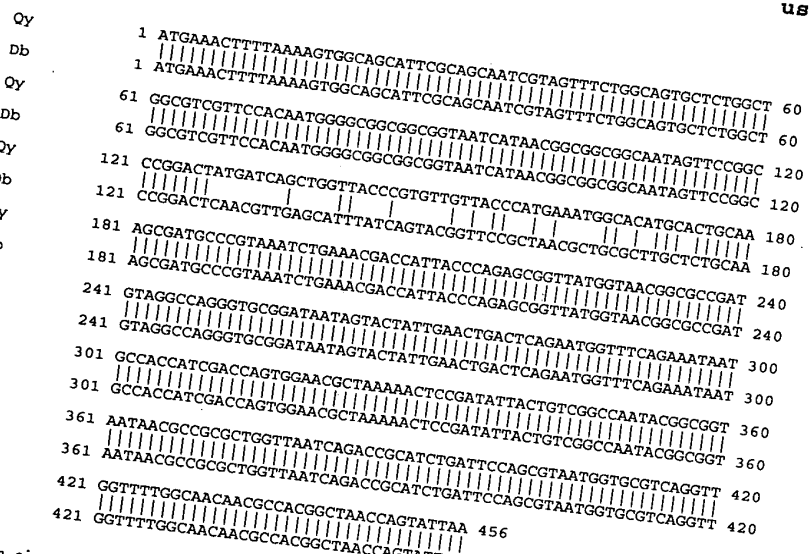
SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	406.4	89.1	456	1	US-08-233-788A-58	Sequence 58, Appl
2	282.2	61.9	361	1	US-08-233-788A-56	Sequence 56, Appl
c 3	36.4	8.0	798	4	US-09-252-991A-13717	Sequence 13717, A
c 4	36.4	8.0	2748	4	US-09-252-991A-13820	Sequence 13820, A
5	36.4	8.0	3720	4	US-09-252-991A-13627	Sequence 13627, A
6	33.2	7.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
7	33.2	7.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
8	33	7.2	834	4	US-09-621-976-2574	Sequence 2574, Ap
9	32.4	7.1	332	2	US-08-469-412A-8	Sequence 8, Appli
10	32.4	7.1	332	3	US-09-021-715-8	Sequence 8, Appli
c 11	31.4	6.9	2007	4	US-09-643-597-153	Sequence 153, App
c 12	31.4	6.9	2007	4	US-09-480-884A-153	Sequence 153, App
c 13	31.4	6.9	2007	4	US-09-542-615A-153	Sequence 153, App
c 14	31.4	6.9	2007	4	US-09-606-421B-153	Sequence 153, App
c 15	31.4	6.9	2007	4	US-09-221-107-153	Sequence 153, App
c 16	31.4	6.9	2148	4	US-09-643-597-154	Sequence 154, App
c 17	31.4	6.9	2148	4	US-09-480-884A-154	Sequence 154, App
c 18	31.4	6.9	2148	4	US-09-542-615A-154	Sequence 154, App
c 19	31.4	6.9	2148	4	US-09-606-421B-154	Sequence 154, App
c 20	31.4	6.9	2148	4	US-09-221-107-154	Sequence 154, App
c 21	31.2	6.8	1212	4	US-09-252-991A-7452	Sequence 7452, Ap
22	31.2	6.8	1425	4	US-09-252-991A-7188	Sequence 7188, Ap
c 23	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c 24	31.2	6.8	1692	4	US-09-252-991A-7513	Sequence 7513, Ap
c 25	31.2	6.8	2436	4	US-09-540-236-492	Sequence 492, App
c 26	31.2	6.8	65792	4	US-09-596-002-31	Sequence 31, Appl
27	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
28	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c 29	30.4	6.7	552	4	US-09-252-991A-9800	Sequence 9800, Ap
c 30	30.4	6.7	894	4	US-09-434-354-1	Sequence 1, Appli
c 31	30.2	6.6	774	4	US-09-252-991A-11893	Sequence 11893, A
c 32	30.2	6.6	1083	4	US-09-252-991A-11930	Sequence 11930, A
33	30.2	6.6	3323	2	US-08-422-699A-10	Sequence 10, Appl
34	30.2	6.6	3323	2	US-08-422-706B-10	Sequence 10, Appl
c 35	29.8	6.5	1500	4	US-09-252-991A-6254	Sequence 6254, Ap
36	29.8	6.5	1695	4	US-09-252-991A-6044	Sequence 6044, Ap
c 37	29.8	6.5	1747	4	US-09-566-921-66	Sequence 66, Appl
c 38	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
c 39	29.4	6.4	832	4	US-09-621-976-2813	Sequence 2813, Ap
c 40	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c 41	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c 42	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
43	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c 44	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, App
c 45	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, App

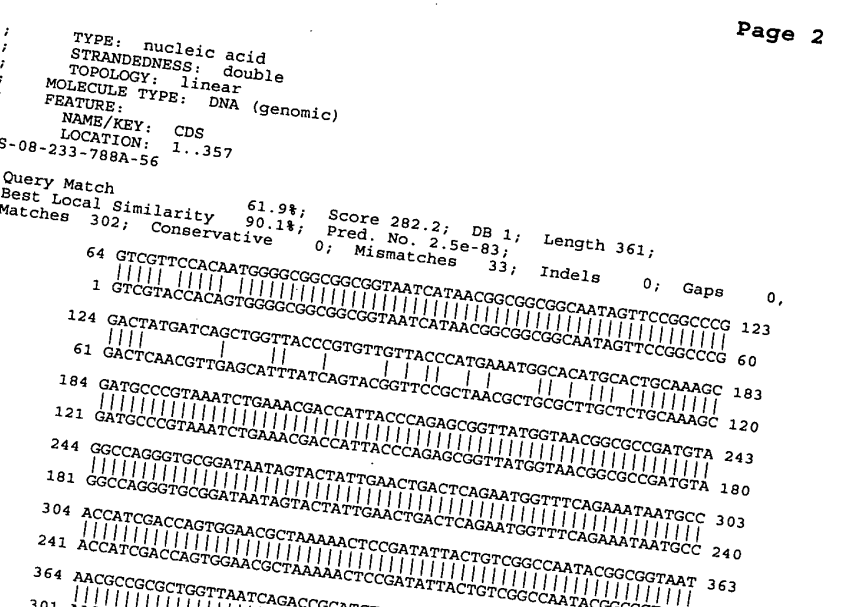
ALIGNMENTS

```
RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-788A-58

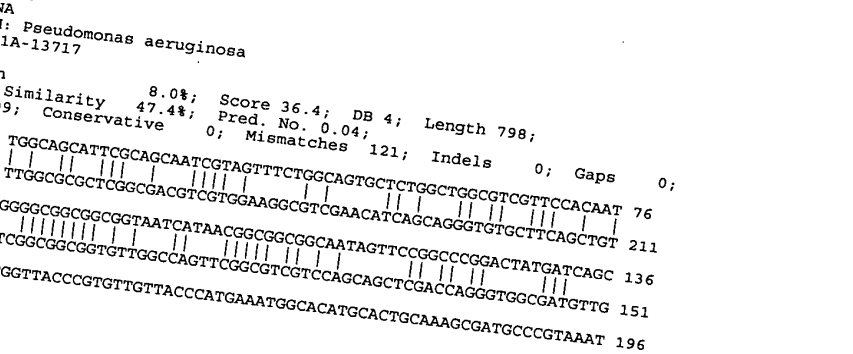
  Query Match            89.1%;  Score 406.4;  DB 1;  Length 456;
  Best Local Similarity  93.2%;  Pred. No. 3.1e-124;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;
```



RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56



RESULT 3
US-09-252-991A-13717/c
; Sequence 13717, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13717
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13717



```
Db        150 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 91

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
              ||      || ||  | |  |||| | |  | |||| || ||||||
Db         90 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAAACCGGCCTTGACGTAGGC 41


RESULT 4
US-09-252-991A-13820/c
; Sequence 13820, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 13820
;   LENGTH: 2748
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13820

  Query Match             8.0%;  Score 36.4;  DB 4;  Length 2748;
  Best Local Similarity   47.4%;  Pred. No. 0.074;
  Matches  109;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | |  ||  |||  |   |||| |     | |     || |   || ||  |||  | |
Db        548 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 489

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              |||||||| | |   ||   ||||| || | |       || || ||      |||
Db        488 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGCTCGACCAGGGTGGCGATGTTG 429

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
              | |   |||    || | ||||  | | |       | |||    ||||    |   |
Db        428 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 369

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
              ||      || ||  | |  |||| | |  | |||| || ||||||
Db        368 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAAACCGGCCTTGACGTAGGC 319


RESULT 5
US-09-252-991A-13627
; Sequence 13627, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 13627
;   LENGTH: 3720
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13627

  Query Match             8.0%;  Score 36.4;  DB 4;  Length 3720;
  Best Local Similarity   47.4%;  Pred. No. 0.087;
  Matches  109;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | |  ||  |||  |   |||| |     | |     || |   || ||  |||  | |
Db       3080 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 3139

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              |||||||| | |   ||   ||||| || | |       || || ||      |||
Db       3140 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGCTCGACCAGGGTGGCGATGTTG 3199

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
              | |   |||    || | ||||  | | |       | |||    ||||    |   |
Db       3200 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 3259

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
              ||      || ||  | |  |||| | |  | |||| || ||||||
Db       3260 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAAACCGGCCTTGACGTAGGC 3309


RESULT 6
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             7.3%;  Score 33.2;  DB 3;  Length 4403765;
  Best Local Similarity   44.6%;  Pred. No. 31;
  Matches  131;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              | || | ||  |||| ||       | ||| | |  |   |   || | || | || | |  |
Db    1093443 GCGGTGCCGACGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGTCGGGTCGATCACCGAC 1093502

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
                 |   ||| |       | |    | || ||    | | | || ||| |
Db    1093503 ACCGTGTTGCTGTCGAAGTTGGTCACGAAGACCAGGCCGGTGACGGGGCTGACCGCCACC 1093562

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGG 256
              | |    |||| || || |  | | |   ||   ||| | || ||| ||    || | |
Db    1093563 CCGCTTGGACCGTTGCCGATGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGCGGGGTTG 1093622

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316
              || |  |      ||      | | |||| |  | |   |||| ||   |  |
Db    1093623 ATCACCGACACCGTGCCGCTGCCGAAATTGGTGACGTAGACGTTGCCGCCCGGGTTGACC 1093682

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              |  || |        || |    ||| ||      | || ||| |   |||||
Db    1093683 GCCACCCCGTGCGGATCGTTGAAGCTGGCGTGGGTGATGGTGGTGACGGCGCCG 1093736


RESULT 7
```

```
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             7.3%;  Score 33.2;  DB 3;  Length 4411529;
  Best Local Similarity   44.6%;  Pred. No. 31;
  Matches  131;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              | ||| | || |||| ||      | ||| | |  |  |   || | || | || | |  |
Db    1093444 GCGGTGCCGACGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGTCGGGTCGATCACCGAC 1093503

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
                |   ||| |     | |    |  || ||     |  | | || |||  |
Db    1093504 ACCGTGTTGCTGTCGAAGTTGGTCACGAAGACCAGGCCGGTGACGGGGCTGACCGCCACC 1093563

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGG 256
              | |    |||| || || |  | | |   ||   ||| | || ||| ||    || |  |
Db    1093564 CCGCTTGGACCGTTGCCGATGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGCGGGGTTG 1093623

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316
              || |  |       ||      |  | | |||  |  | |   |||| ||   |   |
Db    1093624 ATCACCGACACCGTGCCGCTGCCGAAATTGGTGACGTAGACGTTGCCGCCCGGGTTGACC 1093683

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              |  || |         || |    ||| ||       | || ||| |   |||||
Db    1093684 GCCACCCCGTGCGGATCGTTGAAGCTGGCGTGGGTGATGGTGGTGACGGCGCCG 1093737


RESULT 8
US-09-621-976-2574
; Sequence 2574, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2574
;   LENGTH: 834
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 497..826
US-09-621-976-2574

  Query Match             7.2%;  Score 33;  DB 4;  Length 834;
  Best Local Similarity   14.9%;  Pred. No. 0.54;
  Matches   18;  Conservative   64;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         79 GGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTG 138
              :|  |: :: : :: :      ::| : :| |  :::| :: |: |::: : :::   ::
Db         35 RGGWGSCRSTSRMRRSKSYGARRCMSRRGKAKCMSWTGMSWCSTGRMSSWRKRKRSWSYW 94

Qy        139 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCT 198
              :|:: : |:::|  ::::  :::::: ::: :::::  |:::|:|:  |::|: |::|:
Db         95 STKWKMRGYSWTYSYRMYRCYSMWWWMCMRSCYKSRSRMCMMWGSGRSMCMYGWSARWCW 154

Qy        199 G 199
              :
Db        155 R 155


RESULT 9
US-08-469-412A-8
; Sequence 8, Application US/08469412A
; Patent No. 5856125
;  GENERAL INFORMATION:
;    APPLICANT:  Mavrothalassitis, George J.
;    APPLICANT:  Blair, Donald G.
;    APPLICANT:  Fisher, Robert J.
;    APPLICANT:  Beal Jr., Gregory J.
;    APPLICANT:  Athanasiou, Meropi A.
;    APPLICANT:  Sgouras, Dionyssios N.
;    TITLE OF INVENTION:  The ERF Genetic Locus and Its Products
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,412A
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  015280-229000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  332 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..332
;      OTHER INFORMATION:  /note= "human ERF gene promoter region"
US-08-469-412A-8

  Query Match             7.1%;  Score 32.4;  DB 2;  Length 332;
  Best Local Similarity   47.1%;  Pred. No. 0.53;
  Matches   99;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy         47 GCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCG 106
              || | |    ||| ||||  ||  | | |||| |  |||||||| |       |||| | |
Db        120 GCGGGGAGGGGGCGGGCGCAGTGTCTCCATGGCGACGCGGCGGTGACGTCGCCGGCCGGG 179

Qy        107 GCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGG 166
              |     |  |  || |  |     |  | |  | |||||  |  |   |    |  | ||
```

```
Db        180 GGGCGTGGGCGTCCCGGCCCCGGAGTGCGATATTAACCCGGGAGGCGGCGGCGGGGAGGG 239

Qy        167 CACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATG 226

Db        240 GAGAGGCTCTGAGAGGCGAGGCCGGGTGAGGCGGCGAGGGCGGCCCGACGGGCGCGGGAC 299

Qy        227 GTAACGGCGCCGATGTAGGCCAGGGTGCGG 256

Db        300 GGGACGGGGCAGCGAGGGCGCCGGGAGCCG 329


RESULT 10
US-09-021-715-8
; Sequence 8, Application US/09021715
; Patent No. 6194547
;   GENERAL INFORMATION:
;        APPLICANT: Mavrothalassitis, George J.
;                   Blair, Donald G.
;                   Fisher, Robert J.
;                   Beal Jr., Gregory J.
;                   Athanasiou, Meropi A.
;                   Sgouras, Dionyssios N.
;        TITLE OF INVENTION: The ERF Genetic Locus and Its Products
;        NUMBER OF SEQUENCES: 16
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/021,715
;             FILING DATE: 10-Feb-1998
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Garrett-Wackowski, Eugenia
;             REGISTRATION NUMBER: 37,330
;             REFERENCE/DOCKET NUMBER: 015280-229000
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 576-0200
;             TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 332 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  -
;             LOCATION:  1..332
;             OTHER INFORMATION:  /note= "human ERF gene promoter region"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-021-715-8

  Query Match             7.1%;  Score 32.4;  DB 3;  Length 332;
  Best Local Similarity   47.1%;  Pred. No. 0.53;
  Matches   99;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy         47 GCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCG 106

Db        120 GCGGGGAGGGGGCGGGCGCAGTGTCTCCATGGCGACGCGGCGGTGACGTCGCCGGCCGGG 179

Qy        107 GCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGG 166

Db        180 GGGCGTGGGCGTCCCGGCCCCGGAGTGCGATATTAACCCGGGAGGCGGCGGCGGGGAGGG 239

Qy        167 CACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATG 226

Db        240 GAGAGGCTCTGAGAGGCGAGGCCGGGTGAGGCGGCGAGGGCGGCCCGACGGGCGCGGGAC 299

Qy        227 GTAACGGCGCCGATGTAGGCCAGGGTGCGG 256

Db        300 GGGACGGGGCAGCGAGGGCGCCGGGAGCCG 329


RESULT 11
US-09-643-597-153/c
; Sequence 153, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.455C11
;  CURRENT APPLICATION NUMBER: US/09/643,597
;  CURRENT FILING DATE:  2000-08-21
;  NUMBER OF SEQ ID NOS:  369
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-643-597-153

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy         119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178

Db        1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy         179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238

Db        1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy         239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298

Db        1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy         299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335

Db        1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 12
US-09-480-884A-153/c
; Sequence 153, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C6
;  CURRENT APPLICATION NUMBER: US/09/480,884A
```

```
;  CURRENT FILING DATE:  2001-08-27
;  NUMBER OF SEQ ID NOS: 330
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-480-884A-153

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||  |  |  ||| || |  ||| |||      ||  | |  ||   |  |||    |
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              |     | |  | ||||  |  |  |   |||   ||  || |||  |  ||| |   |
Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              ||||  |  |  | ||  |   |    |||  || ||| |    | |    |||
Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              ||    ||  | |  ||||| | |  | || ||||
Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 13
US-09-542-615A-153/c
; Sequence 153, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Fanger, Gary R.
;  TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.455C8
;  CURRENT APPLICATION NUMBER: US/09/542,615A
;  CURRENT FILING DATE:  2000-04-14
;  NUMBER OF SEQ ID NOS:  350
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-542-615A-153

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||  |  |  ||| || |  ||| |||      ||  | |  ||   |  |||    |
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              |     | |  | ||||  |  |  |   |||   ||  || |||  |  ||| |   |
Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              ||||  |  |  | ||  |   |    |||  || ||| |    | |    |||
Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              ||    ||  | |  ||||| | |  | || ||||
Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 14
US-09-606-421B-153/c
; Sequence 153, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C9
;  CURRENT APPLICATION NUMBER: US/09/606,421B
;  CURRENT FILING DATE:  2000-06-28
;  NUMBER OF SEQ ID NOS: 358
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-606-421B-153

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||  |  |  ||| || |  ||| |||      ||  | |  ||   |  |||    |
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              |     | |  | ||||  |  |  |   |||   ||  || |||  |  ||| |   |
Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              ||||  |  |  | ||  |   |    |||  || ||| |    | |    |||
Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              ||    ||  | |  ||||| | |  | || ||||
Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 15
US-09-221-107-153/c
; Sequence 153, Application US/09221107
; Patent No. 6660838
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
;  FILE REFERENCE: 210121.455C2
;  CURRENT APPLICATION NUMBER: US/09/221,107
;  CURRENT FILING DATE: 1998-12-22
;  NUMBER OF SEQ ID NOS: 161
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Human
US-09-221-107-153

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
```

```
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238

Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298

Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299.ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335

Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221
```

Search completed: March 18, 2004, 20:20:48
Job time : 64.4971 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	36.4	8.0	2610	12	US-10-282-122A-30140	Sequence 30140, A
c 2	36	7.9	2667	12	US-10-282-122A-33185	Sequence 33185, A
c 3	35.8	7.9	954	12	US-10-282-122A-13780	Sequence 13780, A
4	34.2	7.5	1355	15	US-10-045-674-593	Sequence 593, App
5	34.2	7.5	2547	12	US-10-282-122A-36154	Sequence 36154, A
c 6	34	7.5	1878	15	US-10-369-493-43934	Sequence 43934, A
7	34	7.5	2886	9	US-09-801-368-131	Sequence 131, App
8	33.6	7.4	2301	9	US-09-815-242-4844	Sequence 4844, Ap
9	33.6	7.4	2301	9	US-09-815-242-9042	Sequence 9042, Ap
10	33.4	7.3	3140	12	US-10-425-114-29450	Sequence 29450, A
11	33.4	7.3	3485	12	US-10-424-599-64614	Sequence 64614, A
c 12	33.2	7.3	493	12	US-10-282-122A-26087	Sequence 26087, A
c 13	33.2	7.3	992	12	US-10-425-114-31747	Sequence 31747, A
14	32.8	7.2	568	14	US-10-333-631-1	Sequence 1, Appli
15	32.8	7.2	718	15	US-10-027-632-26874	Sequence 26874, A
16	32.8	7.2	1059	9	US-09-738-626-482	Sequence 482, App
c 17	32.8	7.2	1113	9	US-09-738-626-481	Sequence 481, App
c 18	32.8	7.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
c 19	32.6	7.1	960	12	US-10-282-122A-11889	Sequence 11889, A
c 20	32.4	7.1	789	14	US-10-156-761-2022	Sequence 2022, Ap
c 21	32.4	7.1	9025608	14	US-10-156-761-1	Sequence 1, Appli
22	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
23	32.2	7.1	1015	14	US-10-198-846-4377	Sequence 4377, Ap
c 24	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
c 25	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
26	32	7.0	669	12	US-10-282-122A-31444	Sequence 31444, A
c 27	32	7.0	1001	12	US-10-425-114-14270	Sequence 14270, A
c 28	32	7.0	1152	14	US-10-156-761-303	Sequence 303, App
29	32	7.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
c 30	31.8	7.0	678	14	US-10-156-761-1219	Sequence 1219, Ap
31	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
32	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
33	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c 34	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
c 35	31.8	7.0	2766	12	US-10-282-122A-42042	Sequence 42042, A
36	31.6	6.9	951	15	US-10-369-493-43201	Sequence 43201, A
c 37	31.6	6.9	1338	15	US-10-369-493-37494	Sequence 37494, A
c 38	31.6	6.9	1642	9	US-09-887-576-469	Sequence 469, App
39	31.4	6.9	734	12	US-10-425-114-566	Sequence 566, App
c 40	31.4	6.9	1320	10	US-09-738-630-70	Sequence 70, Appl
c 41	31.4	6.9	1467	12	US-10-282-122A-25701	Sequence 25701, A
c 42	31.4	6.9	1833	9	US-09-989-920-87	Sequence 87, Appl
c 43	31.4	6.9	2007	9	US-09-735-705-153	Sequence 153, App
c 44	31.4	6.9	2007	9	US-09-850-716A-153	Sequence 153, App
c 45	31.4	6.9	2007	9	US-09-897-778-153	Sequence 153, App

ALIGNMENTS

RESULT 1
US-10-282-122A-30140/c
; Sequence 30140, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

```
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30140
;    LENGTH: 2610
;    TYPE: DNA
;    ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30140

  Query Match
  Best Local Similarity   8.0%;  Score 36.4;  DB 12;  Length 2610;
  Matches  109;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy          17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
Db         410 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 351

Qy          77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
Db         350 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGTTCGACCAGGGTGGCGATGTTG 291

Qy         137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
Db         290 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGAG 231

Qy         197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
Db         230 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAATCCGGCCTTGACGTAGGC 181

RESULT 2
US-10-282-122A-33185/c
; Sequence 33185, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33185
;    LENGTH: 2667
;    TYPE: DNA
;    ORGANISM: Pseudomonas syringae
US-10-282-122A-33185

  Query Match
  Best Local Similarity   7.9%;  Score 36;  DB 12;  Length 2667;
  Matches  159;  Conservative    0;  Mismatches  205;  Indels    0;  Gaps    0;

Qy          25 TTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGC 84
Db         471 TTCGGCGACGTCGTGGAAGGCATCAAACATCAGCAGAGTGTGCTTGAGTTGTTCGGCGGC 412

Qy          85 GGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACC 144
Db         411 GACCGGGGCCAGCGTGGCGTCGTCGAGCAGTTCGACCAGTGTGACGATGTTGTAGCCGCC 352

Qy         145 CGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACG 204
Db         351 CTGCATGGTGCCCAGCAGCTCGGTTGCGCGCTTTTTGTCGATCAGGGGAGAAGTGGCTTC 292

Qy         205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
Db         291 GCCCTTGGCCAGGGCAGACAGGGAACCGGCCTTGACGTAGGCAGCTTCGTCAACACCCGG 232

Qy         265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
Db         231 TGGAATGCGATTGGTGATCAGGTCGACAAGGAAAGCTTCTTCGCCAGCCGGGGGATTTTT 172

Qy         325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAG 384
Db         171 CAGCAGCTCGACCAGGCCTGCGGTTTGTTCGGCGTTAAGCGGCTGGGGCACGATACCCTG 112

Qy         385 ACCG 388
Db         111 GGCG 108

RESULT 3
US-10-282-122A-13780/c
; Sequence 13780, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER  60/230,347
```

```
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13780
;   LENGTH: 954
;   TYPE: DNA
;   ORGANISM: Burkholderia mallei
US-10-282-122A-13780

  Query Match             7.9%;  Score 35.8;  DB 12;  Length 954;
  Best Local Similarity   54.1%;  Pred. No. 0.2;
  Matches   73;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy         55 CTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGT 114
              ||||| || | |     | | ||| |||||||||| |||     ||||||  ||      |
Db        300 CTGGCGCGCATAGCCGAAGTAGGGGATGGCGGCGGTGATCCGGCCGGCGGATGCGCGCTT 241

Qy        115 TCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 174
                ||   ||| ||||||| | | |   |  |||||    | |      |||| | |
Db        240 GAGCGCATCGACCATGATCATCAGTTCCATCAGGTTGTCGTTCGTCGGCGCGCACGTCGA 181

Qy        175 CTGCAAAGCGATGCC 189
                ||||   ||| | |
Db        180 TTGCAGGACGAAGAC 166


RESULT 4
US-10-045-674-593
; Sequence 593, Application US/10045674
; Publication No. US20030232333A1
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 60/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674-593

  Query Match             7.5%;  Score 34.2;  DB 15;  Length 1355;
  Best Local Similarity   26.9%;  Pred. No. 0.86;
  Matches   90;  Conservative   76;  Mismatches  169;  Indels    0;  Gaps    0;

Qy         20 CAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGG 79
              |: | .|:   ||   :|  | ::::||: |:| :   | :||:| :  :  :    | |
Db        755 CWCCKGTYAAYGCKGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWG 814

Qy         80 GCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGG 139
              |: ::|  ||:  :  :: :|| ||:||:: : |::::||:     : |: |: |:  :
Db        815 GYWSYGARGGWGGYGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARA 874

Qy        140 TTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTG 199
              |  |   :| :  :|     |  ||| |  |          |     |:|| |   |:::||
Db        875 TGGCWAAYGCYAAYAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTG 934

Qy        200 AAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              |  | |    :|  :    : ::   | :|  | :  :| ||  | :   | :|    :|
Db        935 AYGCYAARGGYAARYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYA 994

Qy        260 ATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGA 319
               : |:  : |:         :   ||:||:  : |:  :| :|    : |:|     | :|
Db        995 TYGGYGAYGTYWSYGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKA 1054

Qy        320 ACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              |  |: |:|   |: |  |:  :| :||  |  |:
Db       1055 ATTCYCARATGGCYCARGTYGGWGAYGGKGAYAAY 1089


RESULT 5
US-10-282-122A-36154
; Sequence 36154, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 36154
;   LENGTH: 2547
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
US-10-282-122A-36154

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 2547;
  Best Local Similarity   47.4%;  Pred. No. 1.2;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy        133 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGT 192
              |||| |||| | | ||||||| |||||||| |||| |||  | |   |  ||
Db        850 CAGCAGGTTGCACTTGTTGTTGCCCATGAACTGGCGCATCAATGGTTTGGTAATCTGGTA 909

Qy        193 AAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGT 252
              |       |    |      |  ||   |     |||  |    ||| || |     |  |
Db        910 ACTATGAAGTGGTGGGATGACCTTTGGCTGAATGAAAGTTTTGCCAATATGATGGAATAT 969

Qy        253 GCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGAC 312
              |  | ||  |  | |||||  |||   |    | ||  |||    ||   |   |||| ||
Db        970 GTTGCTATTGATGCTATTAAACCAAACTGGCATATTTTTGAAGACTTTCAGACCAGCGGT 1029

Qy        313 CAGTGGAACGCTAAAAACTCCGATATTACTGTCGG 347
                  | |||   ||    |||   |||| |||
Db       1030 GTTCCTCATGCTTTGAAGCGTGATGCAACTGACGG 1064


RESULT 6
US-10-369-493-43934/c
; Sequence 43934, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 43934
;   LENGTH: 1878
;   TYPE: DNA
;   ORGANISM: No. US20030233675A1toc punctiforme
US-10-369-493-43934

  Query Match             7.5%;  Score 34;  DB 15;  Length 1878;
  Best Local Similarity   46.0%;  Pred. No. 1.2;
  Matches  115;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy         96 TAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTAC 155
              ||| |||| ||| |||| ||| || |  |  |  |  | | |  || |||||
Db        553 TAAAGGCGACGGAAATAACTCCTTCCTTGTGTCCGCGCCAAGGTTGAGTAATGGCGTTAC 494

Qy        156 CCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCA 215
              |    ||||  |||| |||  |  ||     ||  |||| |  |      ||
Db        493 CTTGTAAATCCCACAAACACACCGTTCCGTCAAAACCAACACTGACAATGCTTACCCCAG 434

Qy        216 GAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACT 275
              | | |  |   |  |||   |   |  | || |  ||||   |  |      |
Db        433 AGGGGTTGGCGCAGCCATCGCTATTTGGGCTAAAGGCGACGGAAATAACTCCTTCCTCAT 374

Qy        276 GACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              | | | |    | || || ||| |   ||| | |          |  | |||| ||||
Db        373 GTCCCTGCCACGGTTGAGTAATGAGATTACCTTGTAGATCCCATAAACACAAAATTCCGT 314

Qy        336 TATTACTGTC 345
                | |||||


Db        313 CAATACTGCC 304


RESULT 7
US-09-801-368-131
; Sequence 131, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 131
;   LENGTH: 2886
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-131

  Query Match             7.5%;  Score 34;  DB 9;  Length 2886;
  Best Local Similarity   50.6%;  Pred. No. 1.5;
  Matches   82;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              | ||  | | | | || | | | |||||     | ||     || |    |    |||| | |
Db       1501 AACAACAACAATAACGATAACGATAACGATAACAATAATAGCAATAATAATAATAACAAT 1560

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || | |     | | | ||| |  ||  ||||||  ||  | | ||  |    ||| |
Db       1561 AATAATAATAACAATAATAATAATAACAATAATAATAACAATAACAACAACAACAACAAT 1620

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
              || ||     ||| || |      || |  || | |||||
Db       1621 AATAATAATAATAGTAATAATATTAAAAATAACGTTGATAAC 1662


RESULT 8
US-09-815-242-4844
; Sequence 4844, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
```

```
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4844
;   LENGTH: 2301
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-815-242-4844

  Query Match             7.4%;  Score 33.6;  DB 9;  Length 2301;
  Best Local Similarity   47.6%;  Pred. No. 1.9;
  Matches   99;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy        159 TGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAG 218
              |||||  | | |||||  ||      |||| | | |||| ||||       | |     | |
Db       1146 TGAAAGCGTATATGCAAAGCTTGATGATGTCAGCAAATTTGAAGATGTGACTGATAATAT 1205

Qy        219 CGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGAC 278
              |||  | |    |   |  |||  ||  |    |    |   || || | | || |  |
Db       1206 GAGTTTAGATTTTGATACTAATGGTGGTTATTCTTTAAACTTTAATAATTTAGACCAAAG 1265

Qy        279 TCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATAT 338
              | | |||  | | | |||  |||          |  | |  || ||||  ||||  ||  |
Db       1266 TAAAAATTATGTAATAAAATATGAAGGGTATTATGATTCAAATGCTAGCAACTTAGAATT 1325

Qy        339 TACTGTCGGCCAATACGGCGGTAATAAC 366
              |        ||  |  ||   | |||||
Db       1326 TCAAACACACCTTTTTGGATATTATAAC 1353


RESULT 9
US-09-815-242-9042
; Sequence 9042, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9042
;   LENGTH: 2301
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2301)
US-09-815-242-9042

  Query Match             7.4%;  Score 33.6;  DB 9;  Length 2301;
  Best Local Similarity   47.6%;  Pred. No. 1.9;
  Matches   99;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy        159 TGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAG 218
              |||||  | | |||||  ||      |||| | | |||| ||||       | |     | |
Db       1146 TGAAAGCGTATATGCAAAGCTTGATGATGTCAGCAAATTTGAAGATGTGACTGATAATAT 1205

Qy        219 CGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGAC 278
              |||  | |    |   |  |||  ||  |    |    |   || || | | || |  |
Db       1206 GAGTTTAGATTTTGATACTAATGGTGGTTATTCTTTAAACTTTAATAATTTAGACCAAAG 1265

Qy        279 TCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATAT 338
              | | |||  | | | |||  |||          |  | |  || ||||  ||||  ||  |
Db       1266 TAAAAATTATGTAATAAAATATGAAGGGTATTATGATTCAAATGCTAGCAACTTAGAATT 1325

Qy        339 TACTGTCGGCCAATACGGCGGTAATAAC 366
              |        ||  |  ||   | |||||
Db       1326 TCAAACACACCTTTTTGGATATTATAAC 1353


RESULT 10
US-10-425-114-29450
; Sequence 29450, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 29450
;   LENGTH: 3140
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-GMFLMINSOY044G01_FLI
US-10-425-114-29450

  Query Match             7.3%;  Score 33.4;  DB 12;  Length 3140;
  Best Local Similarity   45.8%;  Pred. No. 2.6;
  Matches  115;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy         91 AATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTT 150
              |||||    |   |||   |||||        | ||   ||||| |      |  |  ||
Db       1801 AATCAAGGCAAGGGCCTAAATAGGCTAAATCAGGTAGATGATAAAGGCATAGCTGGTAAA 1860

Qy        151 GTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATT 210
              | |   || |   ||    | | || | |  ||||| |  ||   ||||| |    |
Db       1861 ACTGCATCTGGAGGAGCTAGAGAAATGGACAAGGATGCTCAGAATGTTGAAACAAGTGGT 1920

Qy        211 ACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATT 270
              |  | ||  ||| ||  || ||      |||| |   |   | |||||| |     ||
Db       1921 TCAGATAGTAGTTCTGCAAAAGGAAAGAATGTTGTTGATAATATGGATAATGGAGAGCTT 1980
```

```
Qy        271 GAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAAC 330
Db       1981 TCAAAATCGAATGAGCGTCTTAAAAGAACTGCAGTTGAGGAAAATCCAGAGGATGAAAAG 2040

Qy        331 TCCGATATTAC 341
Db       2041 ATTGAACTTTC 2051
```

RESULT 11
US-10-424-599-64614
; Sequence 64614, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 64614
; LENGTH: 3485
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_29358C.1
US-10-424-599-64614

Query Match 7.3%; Score 33.4; DB 12; Length 3485;
Best Local Similarity 45.8%; Pred. No. 2.7;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy         91 AATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTT 150
Db       2146 AATCAAGGCAAGGGCCTAAATAGGCTAAATCAGGTAGATGATAAAGGCATACCTGGTAAA 2205

Qy        151 GTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATT 210
Db       2206 ACTGCATCTGGAGGAGCTAGAGAAATGGACAAGGATGCTCAGAATGTTGAAACAAGTGGT 2265

Qy        211 ACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATT 270
Db       2266 TCAGATAGTAGTTCTGCAAAAGGAAAGAATGTTGTTGATAATATGGATAATGGAGAGCTT 2325

Qy        271 GAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAAC 330
Db       2326 TCAAAATCAAATGAGCGTCTTAAAAGAACTGCAGTTGAGGAAAATCCAGAGGATGAAAAG 2385

Qy        331 TCCGATATTAC 341
Db       2386 ATTGAACTTTC 2396
```

RESULT 12
US-10-282-122A-26087/c
; Sequence 26087, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26087
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26087

Query Match 7.3%; Score 33.2; DB 12; Length 493;
Best Local Similarity 44.6%; Pred. No. 1.1;
Matches 131; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

```
Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
Db        408 GCGGTGCCGACGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGTCGGGTCGATCACCGAC 349

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
Db        348 ACCGTGTTGCTGTCGAAGTTGGTCACGAAGACCAGGCCGGTGACGGGGCTGACCGCCACC 289

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGG 256
Db        288 CCGCTTGGACCGTTGCCGATGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGCGGGGTTG 229

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316
Db        228 ATCACCGACACCGTGCCGCTGCCGAAATTGGTGACGTAGACGTTGCCGCCCGGGTTGACC 169

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
Db        168 GCCACCCCGTGCGGATCGTTGAAGCTGGCGTGGGTGATGGTGGTGACGGCGCCG 115
```

RESULT 13
US-10-425-114-31747/c
; Sequence 31747, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 31747
;   LENGTH: 992
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMFLB73212F11_FLI
US-10-425-114-31747

  Query Match             7.3%;  Score 33.2;  DB 12;  Length 992;
  Best Local Similarity   52.1%;  Pred. No. 1.6;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         19 GCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGG 78
              || | |  || | ||  | | | |    ||     | ||| |  ||| | | |    |   |
Db        497 GCTGGAGGCGGATCAGGCCTTGGTGGCGGGTCAGGTCTCGGCGGCCTTGGTGGTCTCGGC 438

Qy         79 GGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTG 138
              ||||||||||||    |       |||||||||  | ||  |||| | |  |  | || |
Db        437 GGCGGCGGCGGCCTTGGCGGTGGCGGCGGCCTTGGTGGCGGCTCAGGATCGGGTCTGGGC 378

Qy        139 GTTACCCGTGTTGTTACCCATG 160
              |  | | |||   ||  |  ||
Db        377 GGCAGCGGTGGCCTTGGCGGTG 356


RESULT 14
US-10-333-631-1
; Sequence 1, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer AG
;  TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
;  TITLE OF INVENTION:  ENZYME
;  FILE REFERENCE: LIO121 Foreign Countries
;  CURRENT APPLICATION NUMBER: US/10/333,631
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US 60/221,917
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/280,112
;  PRIOR FILING DATE: 2001-04-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 568
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90)..(90)
;   OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

  Query Match             7.2%;  Score 32.8;  DB 14;  Length 568;
  Best Local Similarity   51.0%;  Pred. No. 1.6;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        139 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCT 198
              ||  |   || ||    |   ||||   |  | ||| |||||   ||   || ||    ||
Db         87 GTCNCGGCTGCTGCAGCTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCT 146

Qy        199 GAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGAT 258
              |||| |  |  |  ||  ||  |   || || | ||  | | | | ||| ||||||||
Db        147 GAAATGGTCCGTCACCCAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTA 206

Qy        259 AATAGTACTATTGAACTGACTCAGAATGG 287
                | |  ||   ||     | ||   |||
Db        207 CCTGGATCTGCGGATAGCCCACATGCTGG 235


RESULT 15
US-10-027-632-26874
; Sequence 26874, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26874
;   LENGTH: 718
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-26874

  Query Match             7.2%;  Score 32.8;  DB 15;  Length 718;
  Best Local Similarity   49.4%;  Pred. No. 1.9;
  Matches   85;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         18 GGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATG 77
              || ||||| | |||||   |  |   || ||  |  | || | || || | | | ||  |
Db        514 GGAAGCATCCACAGCACAAGGTGCCGCTCGCCCTCGTGTGCCCGGGGTGGCTGCCCAGGG 573

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCT 137
              |   | |         | |  ||| |  |||   |||||||   |         |
Db        574 GCTGGCAGCGTCCCCACGTGTCGGGGCTGGCCTGAGTTCCGTGTCCCTGGGATCTACAGC 633

Qy        138 GGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCC 189
              ||   || ||  |  |  |||||   |||||| |||| |     | |||||
Db        634 TGTGTTCCCTGAAGGCAGACATGACTGGGCACAGGCACGGGGGTTCTATGCC 685


Search completed: March 19, 2004, 03:49:54
Job time : 264.734 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-17	Sequence 17, Appl
2	406.4	89.1	456	6	US-08-233-642A-56	Sequence 56, Appl
3	406.4	89.1	456	23	US-09-543-407-1	Sequence 1, Appli
4	363.2	79.6	456	23	US-09-543-407-19	Sequence 19, Appl
5	355.2	77.9	456	23	US-09-543-407-11	Sequence 11, Appl
6	355.2	77.9	456	23	US-09-543-407-13	Sequence 13, Appl
7	355.2	77.9	456	23	US-09-543-407-23	Sequence 23, Appl
8	355.2	77.9	456	23	US-09-543-407-25	Sequence 25, Appl
9	353.6	77.5	456	23	US-09-543-407-15	Sequence 15, Appl
10	352	77.2	456	23	US-09-543-407-21	Sequence 21, Appl
11	352	77.2	456	23	US-09-543-407-27	Sequence 27, Appl
12	348.8	76.5	456	23	US-09-543-407-29	Sequence 29, Appl
13	282.2	61.9	361	6	US-08-233-642A-54	Sequence 54, Appl
14	230.4	50.5	456	23	US-09-543-407-3	Sequence 3, Appli
15	230.4	50.5	675	13	US-08-978-878-3	Sequence 3, Appli
16	230.4	50.5	675	31	US-09-741-873B-3	Sequence 3, Appli
17	228	50.0	453	92	US-60-352-946-1	Sequence 1, Appli
18	228	50.0	453	101	US-60-444-371-1	Sequence 1, Appli
19	219.2	48.1	477	17	US-09-252-691-172	Sequence 172, App
20	219.2	48.1	477	17	US-09-252-691C-172	Sequence 172, App
21	219.2	48.1	477	51	US-10-417-886-172	Sequence 172, App
22	186.4	40.9	396	13	US-08-978-878-1	Sequence 1, Appli
23	161.6	35.4	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
c 28	50	11.0	78	23	US-09-543-407-52	Sequence 52, Appl
29	50	11.0	78	23	US-09-543-407-53	Sequence 53, Appl
30	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
31	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
32	41.2	9.0	78	23	US-09-543-407-49	Sequence 49, Appl
33	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
c 34	40.6	8.9	78	23	US-09-543-407-48	Sequence 48, Appl
c 35	39.4	8.6	4762	62	US-60-068-138-239	Sequence 239, App
36	39.4	8.6	47475	33	US-09-806-866A-13	Sequence 13, Appl
37	39.4	8.6	2242716	33	US-09-806-866A-1068	Sequence 1068, Ap
38	39.4	8.6	2272325	43	US-10-018-470A-1	Sequence 1, Appli
39	39.2	8.6	1083	1	PCT-US01-08631-12549	Sequence 12549, A
c 40	37	8.1	14652	33	US-09-806-866A-30	Sequence 30, Appl
c 41	37	8.1	2242716	33	US-09-806-866A-1068	Sequence 1068, Ap
c 42	37	8.1	2272325	43	US-10-018-470A-1	Sequence 1, Appli
43	36.4	8.0	561	17	US-09-252-691-171	Sequence 171, App
44	36.4	8.0	561	17	US-09-252-691C-171	Sequence 171, App
45	36.4	8.0	561	51	US-10-417-886-171	Sequence 171, App

ALIGNMENTS

```
RESULT 1
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-133;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             89.1%;  Score 406.4;  DB 6;  Length 456;
  Best Local Similarity   93.2%;  Pred. No. 1.2e-117;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             89.1%;  Score 406.4;  DB 23;  Length 456;
  Best Local Similarity   93.2%;  Pred. No. 1.2e-117;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             79.6%;  Score 363.2;  DB 23;  Length 456;
  Best Local Similarity   87.3%;  Pred. No. 6e-104;
  Matches  398;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.1e-101;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.1e-101;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.1e-101;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |    | | ||  | |    || | |||  ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||  |||      ||  ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.1e-101;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |    | | ||  | |    || | |      ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||      |           ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 6.7e-101;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |  ||  ||  |||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||  |  ||  |  |||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.2e-100;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||  |  ||  |  |||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.2e-100;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 76.5%; Score 348.8; DB 23; Length 456;
Best Local Similarity 85.3%; Pred. No. 2.2e-99;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 61.9%; Score 282.2; DB 6; Length 361;
Best Local Similarity 90.1%; Pred. No. 2.8e-78;
Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

```
RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-3
```

```
  Query Match             50.5%;  Score 230.4;  DB 23;  Length 456;
  Best Local Similarity   69.1%;  Pred. No. 8.2e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |   ||  ||     | |  | | |    |||| ||| ||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  ||||||||||| |||||   ||| |||||||||   ||     ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||| ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  ||| |||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

```
RESULT 15
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
;  APPLICANT: NORMARK, Staffan
;  APPLICANT: OLSEN, Arne
;  TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
;  FILE REFERENCE: 012889-081
;  CURRENT APPLICATION NUMBER: US/08/978,878
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: SE 8801723-1
;  EARLIER FILING DATE: 1988-05-06
;  EARLIER APPLICATION NUMBER: US 07/347,189
;  EARLIER FILING DATE: 1989-05-04
;  EARLIER APPLICATION NUMBER: US 07/789,437
;  EARLIER FILING DATE: 1991-11-06
;  EARLIER APPLICATION NUMBER: US 07/970,846
;  EARLIER FILING DATE: 1992-11-03
;  EARLIER APPLICATION NUMBER: US 08/187,865
;  EARLIER FILING DATE: 1994-01-28
;  EARLIER APPLICATION NUMBER: US 08/318,519
;  EARLIER FILING DATE: 1994-10-05
;  EARLIER APPLICATION NUMBER: US 08/495,959
;  EARLIER FILING DATE: 1995-06-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-08-978-878-3
```

```
  Query Match             50.5%;  Score 230.4;  DB 13;  Length 675;
  Best Local Similarity   69.1%;  Pred. No. 9.4e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |   ||  ||     | |  | | |    |||| ||| ||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  ||||||||||| |||||   ||| |||||||||   ||     ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||| ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  ||| |||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

Search completed: March 19, 2004, 03:01:56
Job time : 2655.77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	230.4	50.5	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	186.4	40.9	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	34.2	7.5	1355	6	US-10-045-674A-593	Sequence 593, App
c	4	31.4	6.9	435	6	US-10-021-698A-1216	Sequence 1216, Ap
c	5	31.4	6.9	2007	6	US-10-775-972-153	Sequence 153, App
c	6	31.4	6.9	2148	6	US-10-775-972-154	Sequence 154, App
c	7	31.2	6.8	9278	6	US-10-021-698A-796	Sequence 796, App
	8	31	6.8	1099	6	US-10-767-701-8827	Sequence 8827, Ap
	9	31	6.8	150223	6	US-10-624-149A-1	Sequence 1, Appli
c	10	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
c	11	30.4	6.7	883	6	US-10-767-701-1287	Sequence 1287, Ap
c	12	30.2	6.6	989	6	US-10-767-795-6329	Sequence 6329, Ap
c	13	30	6.6	1203	6	US-10-767-701-12486	Sequence 12486, A
c	14	29.8	6.5	1228	6	US-10-767-701-6233	Sequence 6233, Ap
c	15	29.6	6.5	656	6	US-10-767-701-20611	Sequence 20611, A
c	16	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
	17	29.6	6.5	186449	6	US-10-021-698A-715	Sequence 715, App
	18	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
c	19	29.4	6.4	302	6	US-10-767-701-17781	Sequence 17781, A
c	20	29.4	6.4	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	21	29.4	6.4	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	22	29.4	6.4	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	23	29.4	6.4	1400	7	US-60-545-213-6407	Sequence 6407, Ap
	24	29.4	6.4	145597	6	US-10-624-149A-2	Sequence 2, Appli
c	25	29.4	6.4	145597	6	US-10-624-149A-2	Sequence 2, Appli
c	26	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	27	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
	28	29	6.4	594	6	US-10-767-701-15041	Sequence 15041, A
c	29	29	6.4	1002	1	PCT-US04-05654-859	Sequence 859, App
	30	29	6.4	1052	6	US-10-767-701-12607	Sequence 12607, A
	31	29	6.4	2600	6	US-10-451-467A-39	Sequence 39, Appl
	32	28.8	6.3	959	6	US-10-767-701-1446	Sequence 1446, Ap
c	33	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	34	28.6	6.3	592	6	US-10-767-701-22979	Sequence 22979, A
	35	28.6	6.3	749	6	US-10-767-701-12516	Sequence 12516, A
	36	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
c	37	28.4	6.2	270	6	US-10-794-929-95	Sequence 95, Appl
	38	28.4	6.2	378	6	US-10-417-884A-1059	Sequence 1059, Ap
	39	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	40	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
c	41	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	42	28.2	6.2	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	43	28.2	6.2	3416	6	US-10-100-683-3397	Sequence 3397, Ap
	44	28.2	6.2	3776	7	US-60-548-091-34	Sequence 34, Appl
	45	28.2	6.2	3828	7	US-60-548-091-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 50.5%; Score 230.4; DB 5; Length 675;
Best Local Similarity 69.1%; Pred. No. 1.1e-68;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

```
RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873C
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;  LENGTH: 396
;  TYPE: DNA
;  ORGANISM: Escherichia coli
US-09-741-873C-1

  Query Match           40.9%;  Score 186.4;  DB 5;  Length 396;
  Best Local Similarity  66.9%;  Pred. No. 8.4e-54;
  Matches  265;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60
Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

```
RESULT 3
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674A
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 06/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;  LENGTH: 1355
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: M13-III
;  OTHER INFORMATION: nucleotide sequence
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(1305)
US-10-045-674A-593

  Query Match           7.5%;  Score 34.2;  DB 6;  Length 1355;
  Best Local Similarity  26.9%;  Pred. No. 0.098;
  Matches   90;  Conservative   76;  Mismatches  169;  Indels    0;  Gaps    0;

Qy         20 CAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGG 79
Db        755 CWCCKGTYAAYGCKGGYGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWG 814
Qy         80 GCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCCGGACTATGATCAGCTGG 139
Db        815 GYWSYGARGGWGGYGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARA 874
Qy        140 TTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTG 199
Db        875 TGGCWAAYGCYAAYAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTG 934
Qy        200 AAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
Db        935 AYGCYAARGGYAARYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYA 994
Qy        260 ATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGA 319
Db        995 TYGGYGAYGTYWSYGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKA 1054
```

```
Qy        320 ACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              |  |: |:|   |: |  |:  :| :||  |  |:
Db       1055 ATTCYCARATGGCYCARGTYGGWGAYGGKGAYAAY 1089


RESULT 4
US-10-021-698A-1216/c
; Sequence 1216, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;. CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 1216
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-021-698A-1216

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 435;
  Best Local Similarity   46.2%;  Pred. No. 0.51;
  Matches  104;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
              | |  ||   |||||   | ||| | | |    | | | |  || ||||||  | |
Db        353 GATGGTGATGGTGCTGATGGTGGTGATGATAGTGATGGTGATGATGGTGGTAATGGTGAT 294

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 159
              |||   ||  || ||   |     ||   | || |    |  | | ||  || |    ||
Db        293 GGCAATGGTGATGGTGGTGATGGTGATGGTAATGATGGTGGCATCAGTAATGGTGGTGAT 234

Qy        160 GAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
              || | |    ||       | | ||||  || |   |||        | ||  |
Db        233 GATAGTGATGGTGATGATAATGGTGATGATGGTGATGATGATGGTGATAATAGTGATGAT 174

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              ||| ||||| | |  |  |||| ||  |   ||   || || ||
Db        173 GGTAATGGTGATGATGGTGATGATGGTGACCATGTTAATGATGGT 129


RESULT 5
US-10-775-972-153/c
; Sequence 153, Application US/10775972
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C21
;  CURRENT APPLICATION NUMBER: US/10/775,972
;  CURRENT FILING DATE:  2004-02-10
;  NUMBER OF SEQ ID NOS: 563
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-972-153

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 1.1;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||  |  |  ||| || |  ||| |||       ||  | |  ||   |  |||    |
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              |      | |  | |||| |  |  |   |||   ||  || ||| |  ||| |   |
Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              ||||  |   |  | || |  |     |  ||| || ||| |    | |    |||
Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              ||   ||  | |  ||||| | |  | || |||| |
Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 6
US-10-775-972-154/c
; Sequence 154, Application US/10775972
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C21
;  CURRENT APPLICATION NUMBER: US/10/775,972
;  CURRENT FILING DATE:  2004-02-10
;  NUMBER OF SEQ ID NOS: 563
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 154
;   LENGTH: 2148
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-972-154

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 2148;
  Best Local Similarity   46.5%;  Pred. No. 1.1;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||  |  |  ||| || |  ||| |||       ||  | |  ||   |  |||    |
Db       1570 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1511

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              |      | |  | |||| |  |  |   |||   ||  || ||| |  ||| |   |
Db       1510 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1451

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              ||||  |   |  | || |  |     |  ||| || ||| |    | |    |||
Db       1450 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1391

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              ||   ||  | |  ||||| | |  | || |||| |
Db       1390 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1354


RESULT 7
US-10-021-698A-796/c
; Sequence 796, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
```

```
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 796
;    LENGTH: 9278
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-021-698A-796

  Query Match
  Best Local Similarity   6.8%;  Score 31.2;  DB 6;  Length 9278;
  Matches   96;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;
Qy          85 GGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACC 144
Db      347 GGTGATGATGATGGTGACGGTGGTGATGGTGATGGTGATGATGGTGATGATGGTGACGGT 288
Qy         145 CGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACG 204
Db      287 GGTGATGGTGCTGATGATGATGGTGATGCTGATGGTGATGGTGACGGTGATGATGATGGT 228
Qy         205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCCGATGTAGGCCAGGGTGCGGATAATAGT 264
Db      227 GACGGTGATGATGGTGATGATGGTGATGGTGATGCTGATGGTGGTGGTGGTGATGATGGT 168
Qy         265 ACTATTGAACTGACTCAGAATGGT 288
Db      167 GGTGATGATGATGATGATGGT 144

RESULT 8
US-10-767-701-8827
; Sequence 8827, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 8827
;    LENGTH: 1099
;    TYPE: DNA
;    ORGANISM: Sorghum bicolor
;    FEATURE:
;    NAME/KEY: unsure
;    LOCATION: (1)..(1099)
;    OTHER INFORMATION: unsure at all n locations
;    FEATURE:
;    OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS59062_1
US-10-767-701-8827

  Query Match
  Best Local Similarity   6.8%;  Score 31;  DB 6;  Length 1099;
  Matches   76;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;
Qy          99 CGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCA 158
Db      277 CGGCGGCGGCCTCGGTGCGGTTCTTGATCCGGGGGAGCGAGCTCACCTGGGTCGCGCCAA 336
Qy         159 TGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAG 218
Db      337 TGCAGCCGGACCCGAGCGGCGACGCCGGCNATGCGAAGGCGAAGCTGCCGCTGCCGCAGC 396
Qy         219 CGGTTATGGTAACGGCGCCGATGTAGGCCAGG 250
Db      397 CGGTGACGGCTCCGGCGCCGTCGTCGGGGCGG 428

RESULT 9
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
;  APPLICANT: Neubauer, Antonie
;  APPLICANT:  Ziegler, Christina
;  TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
;  FILE REFERENCE: 1/1372
;  CURRENT APPLICATION NUMBER: US/10/624,149A
;  CURRENT FILING DATE:  2003-07-21
;  PRIOR APPLICATION NUMBER: 60/403,282
;  PRIOR FILING DATE: 2002-08-14
;  PRIOR APPLICATION NUMBER: DE 10233064
;  PRIOR FILING DATE: 2002-07-19
;  PRIOR APPLICATION NUMBER: DE 10317008
;  PRIOR FILING DATE: 2003-04-11
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
;    LENGTH: 150223
;    TYPE: DNA
;    ORGANISM: Equine herpesvirus 1
;  PUBLICATION INFORMATION:
;    AUTHORS: Telford,E.A.
;    AUTHORS: Watson, M.S.
;    AUTHORS: McBride, K.
;    AUTHORS: Davison, A.J.
;    TITLE: The DNA sequence of equine herpesvirus-1
;    JOURNAL: Virology
;    VOLUME: 189
;    ISSUE: 1
;    PAGES: 304-316
;    DATE: JUL-1992
;    DATABASE ACCESSION NUMBER: NC 001491, NCBI
;    DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

  Query Match
  Best Local Similarity   6.8%;  Score 31;  DB 6;  Length 150223;
  Matches   73;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;
Qy         176 TGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCG 235
Db      95545 TGGTAACGGTGGGCCTTAGTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACGTCA 95604
Qy         236 CCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
Db      95605 AGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCAGAG 95664
Qy         296 ATAATGCCACCATCGACCAGTGG 318
Db      95665 AGCTTATTCACAACGAGCTGTTG 95687

RESULT 10
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
;  APPLICANT: Bayer Pharmaceuticals Corporation
;  APPLICANT:  Eveleigh, Deepa
;  APPLICANT:  Bigwood, Douglas
;  TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
;  FILE REFERENCE: 5152
;  CURRENT APPLICATION NUMBER: US/10/788,792
;  CURRENT FILING DATE:  2004-02-27
;  PRIOR APPLICATION NUMBER  US 60/450,655
```

```
;  PRIOR FILING DATE: 2003-02-28
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
;   LENGTH: 1575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-788-792-27

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 1575;
  Best Local Similarity   55.7%;  Pred. No. 1.5;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
              |||   | | | | || ||   | | || |  || | |||    || |||| ||||||||
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
              |  | |   | |  ||| |  |   |   |||  |  ||||||||
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583


RESULT 11
US-10-767-701-1287/c
; Sequence 1287, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 1287
;   LENGTH: 883
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(883)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS19621_1
US-10-767-701-1287

  Query Match             6.7%;  Score 30.4;  DB 6;  Length 883;
  Best Local Similarity   61.2%;  Pred. No. 1.6;
  Matches   49;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              | ||||| || | || ||||||      |||  |||   |  ||| ||  | |||||
Db        493 GGTCTGGATGACCTCATTCCACTGCTCAACGGTAGCGTCGAGGATACCGTTGTCGGCAGT 434

Qy        112 AGTTCCGGCCCGGACTATGA 131
              |||||||  | | ||||
Db        433 CTTTCCGGCGTTGGCGATGA 414


RESULT 12
US-10-767-795-6329/c
; Sequence 6329, Application US/10767795
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Zhou, Yihua
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53534)B
;  CURRENT APPLICATION NUMBER: US/10/767,795
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 6329
;   LENGTH: 989
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C1002_1
US-10-767-795-6329

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 989;
  Best Local Similarity   49.7%;  Pred. No. 1.9;
  Matches   77;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy         24 ATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGG 83
              ||   ||| ||  | | |  ||| |  | || ||  |||||| | ||| |         | |
Db        842 ATAAACAGAAACAGCATCTTCAGATAATGATCCTGCTGGCCTTGTGCGGTGCGATGATGA 783

Qy         84 CGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTAC 143
              || |||  |  | | |  ||  | |||     ||||   |  ||    |  ||| |
Db        782 TGGTGGTGGTGGTGATGATGGTATCCATAAGCATGGCCACCAAGATATGAACTTGGAAAG 723

Qy        144 CCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              |||| | |  | |||  |  |||  ||  ||  ||
Db        722 CCGTATGGAAAACCAGCAGCTGGACCAAACATAGC 688


RESULT 13
US-10-767-701-12486/c
; Sequence 12486, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 12486
;   LENGTH: 1203
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS13462_1
US-10-767-701-12486

  Query Match             6.6%;  Score 30;  DB 6;  Length 1203;
  Best Local Similarity   48.3%;  Pred. No. 2.5;
  Matches   84;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy         20 CAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGG 79
              | | |  |||   | ||| || | | | | |||  || | ||  |  ||   |  |   |
Db        271 CCGGAACCGCCTAAGTCGCAGGTGCCGCCGGTGGCCTTGTTGTAGATGTAGTAGGAGTCG 212

Qy         80 GCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGG 139
              ||||  |||   | |  ||||| |||   | ||   |   | | |   |  | || ||
Db        211 CCGGCCCCGGAGGTGAGCGCGGCGACGGTGTTTGTGTTGTAGCAGCCGCCGCCCTGCAGG 152

Qy        140 TTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTA 193
              |   |||| ||  |   |   ||      |||| | ||  || |  |  ||
Db        151 ATGGACGTGCAGTCGGCGCCGCCCTGCGGACCGCACGGGAAGTCGCTCGCTTTA 98


RESULT 14
US-10-767-701-6233/c
; Sequence 6233, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
```

```
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 6233
;   LENGTH: 1228
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS37536_1
US-10-767-701-6233

  Query Match           6.5%;  Score 29.8;  DB 6;  Length 1228;
  Best Local Similarity  52.9%;  Pred. No. 2.9;
  Matches   64;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy         59 CTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCG 118
              ||||||| |     |  | |||  | ||||||| ||  |   ||||| |||  |     | |
Db        666 CTGGCGTAGCCGAAGTAGGGGATGACGGCGGTGATGTTCTTGGCGGAGGCCCTCCGGCAG 607

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||   ||  ||||| ||  | |       ||| |          | ||||  ||  ||||
Db        606 GCGTCGATCATGATGAGGAGCTCCATGAGGTTCTCGTTGGCTGGAGGGCAGGTGGGCTGC 547

Qy        179 A 179
              |
Db        546 A 546


RESULT 15
US-10-767-701-20611/c
; Sequence 20611, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 20611
;   LENGTH: 656
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 11064914
US-10-767-701-20611

  Query Match           6.5%;  Score 29.6;  DB 6;  Length 656;
  Best Local Similarity  52.4%;  Pred. No. 2.5;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy         28 GCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGC 87
              | || | | || |   ||| |  || |  ||  | || || ||||     ||  | || |
Db        126 GTAGAAGTTGTCGAGCCTGCCCTTGTTGAGGTCGTCGGCGATCCATGGGTGGCGGTCGCC 67

Qy         88 GGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGT 147
              ||| | | |      |||||||  |       |||  | ||| | || |  | |     |||
Db         66 GGTGACCTTGTGCACGGCGGCGACGAGCGCGGAGCTGACGAAGAACCCCAGCGACAGCGT 7

Qy        148 GTTG 151
              | ||
Db          6 GCTG 3


Search completed: March 19, 2004, 03:10:15
Job time : 55.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	232.6	51.0	680	12	BJ618688	BJ618688 BJ618688
c	2	37.4	8.2	484	9	AV430994	AV430994 AV430994
	3	37	8.1	590	14	CF848025	CF848025 psMA003xI
	4	37	8.1	610	14	CF842541	CF842541 psHB020xG
c	5	36.6	8.0	891	29	CNS02I89	AL198594 Tetraodon
	6	36.4	8.0	787	28	BZ568375	BZ568375 pacs2-164
	7	35.2	7.7	417	29	FR0020317	AL013202 F.rubripe
c	8	35.2	7.7	811	28	BZ469000	BZ469000 BOOAO58TF
	9	35.2	7.7	813	28	BZ204853	BZ204853 CH230-390
c	10	35	7.7	490	14	CA588919	CA588919 hab51c12.
	11	34.8	7.6	480	9	AU085223	AU085223 AU085223
c	12	34.6	7.6	406	9	AU241511	AU241511 AU241511
c	13	34.6	7.6	499	12	BJ006097	BJ006097 BJ006097
c	14	34.6	7.6	561	12	BJ001420	BJ001420 BJ001420
c	15	34.6	7.6	640	12	BJ000263	BJ000263 BJ000263
c	16	34.6	7.6	656	12	BJ004146	BJ004146 BJ004146
c	17	34.6	7.6	662	12	BJ010653	BJ010653 BJ010653
c	18	34.6	7.6	662	12	BJ011743	BJ011743 BJ011743
c	19	34.6	7.6	671	12	BJ012818	BJ012818 BJ012818
	20	34.6	7.6	682	14	CF402142	CF402142 RTWW1_20_
c	21	34.6	7.6	691	12	BJ000180	BJ000180 BJ000180
	22	34.6	7.6	702	12	BG441624	BG441624 GA__Ea001
	23	34.6	7.6	937	12	BG444255	BG444255 GA__Ea002
c	24	34.4	7.5	365	9	AA746477	AA746477 nw62e02.s
c	25	34.4	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	26	34.4	7.5	827	28	BH438448	BH438448 BOGVQ28TR
	27	34.2	7.5	296	9	AV058630	AV058630 AV058630
	28	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
c	29	34	7.5	588	14	CA731849	CA731849 wlp1c.pk0
c	30	33.8	7.4	522	14	CA707989	CA707989 wdk2c.pk0
	31	33.8	7.4	625	13	BU882267	BU882267 UM74TG08
c	32	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	33	33.6	7.4	584	28	AQ398232	AQ398232 mgxb0001F
	34	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	35	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
c	36	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	37	33.6	7.4	866	29	CNS02C24	AL190597 Tetraodon
	38	33.6	7.4	896	28	CC333807	CC333807 OGMBC13TV
c	39	33.6	7.4	910	14	CA266034	CA266034 SCJLRT307
c	40	33.4	7.3	525	9	AU242732	AU242732 AU242732
c	41	33.4	7.3	596	9	AU177600	AU177600 AU177600
	42	33.4	7.3	649	12	BI261742	BI261742 602954071
	43	33.4	7.3	679	29	CC961526	CC961526 BOIGS10TF
c	44	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
c	45	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 51.0%; Score 232.6; DB 12; Length 680;
Best Local Similarity 69.3%; Pred. No. 4.7e-54;
Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||      | | | | |    |||| ||| ||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||   ||| | |||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
AV430994/c
LOCUS AV430994 484 bp mRNA linear EST 23-AUG-2000
DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone PL027e10_r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
REFERENCE 1 (bases 1 to 484)
AUTHORS Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and Tabata,S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
PUBMED 10907854
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PL027e10_r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 15;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |   | |  | | ||  ||| |||||   ||| ||||| ||||||  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
              | ||  ||   ||  |||| ||| ||||||     |  ||    ||| |  || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323
```

RESULT 3
CF848025
LOCUS CF848025 590 bp mRNA linear EST 30-OCT-2003
DEFINITION psMA003xI14f USDA-IFAFS:Expression of Phytophthora sojae genes during infection and propagation_sMA Phytophthora sojae cDNA clone sMA003I14 5, mRNA sequence.
ACCESSION CF848025
VERSION CF848025.1 GI:38063679
KEYWORDS EST.
SOURCE Phytophthora sojae
ORGANISM Phytophthora sojae
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.
REFERENCE 1 (bases 1 to 590)
AUTHORS Tyler,B.
TITLE Tyler,B. Not Published
JOURNAL Unpublished (2003)
COMMENT Contact: Tyler B
Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtyler@vt.edu
PCR PRimers
FORWARD: BK reverse primer
BACKWARD: BK reverse primer
Plate: 003 row: I column: 14
Seq primer: BK reverse primer
High quality sequence stop: 590.
FEATURES Location/Qualifiers
source 1. .590
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="sMA003I14"
/tissue_type="mycelium"
/cell_line="P6497"
/dev_stage="mycelium"
/lab_host="synthetic medium"
/clone_lib="USDA-IFAFS:Expression of Phytophthora sojae genes during infection and propagation_sMA"
/note="Vector: pBK-CMV; Site_1: EcoR1; Site_2: Xho1"
ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 590;
Best Local Similarity 51.5%; Pred. No. 21;

```
  Matches   85;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy         26 TCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCG 85
              | | |||  | ||  ||  || | |||  || |  |  ||||      ||   | |||||
Db        224 TTGTAGCGGTAGTCCTTGATGTCCGTGAGCTTGGGGAAGTCGCCGATGAAGTAGCCGGCG 283

Qy         86 GCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCC 145
              || | || |  ||| ||||| | | |    | |       | ||||| .  | ||    |
Db        284 CCGTTGATGAGCACGACGGCGCCCACAAAGGCCGTGGCCGCGATGATGGCCAGGCGCTCG 343

Qy        146 GTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCC 190
              | |||   |||  |   ||| |||  |  ||  |||||  ||
Db        344 ATCTTGAAGGCCAGCAGGCCGCAGATGAGCCCCAGCACGATGGCC 388


RESULT 4
CF842541
LOCUS       CF842541                610 bp    mRNA    linear   EST 30-OCT-2003
DEFINITION  psHB020xG21f USDA-IFAFS:Expression of Phytophthora sojae genes
            during infection and propagation_sHB Phytophthora sojae cDNA clone
            sHB020G21 5, mRNA sequence.
ACCESSION   CF842541
VERSION     CF842541.1  GI:38058195
KEYWORDS    EST.
SOURCE      Phytophthora sojae
  ORGANISM  Phytophthora sojae
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Tyler,B.
  TITLE     Tyler,B. Not Published
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Tyler B
            Tyler lab
            VBI
            1880 Pratt Dr., Blacksburg, VA 24061, USA
            Tel: 540-231-7318
            Email: bmtyler@vt.edu
            PCR PRimers
            FORWARD: BK reverse primer
            BACKWARD: BK reverse primer
            Plate: 020  row: G  column: 21
            Seq primer: BK reverse primer
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Phytophthora sojae"
                     /mol_type="mRNA"
                     /db_xref="taxon:67593"
                     /clone="sHB020G21"
                     /tissue_type="mycelium"
                     /cell_line="P6497"
                     /dev_stage="48 hr. post infection stage"
                     /lab_host="Soybean plant"
                     /clone_lib="USDA-IFAFS:Expression of Phytophthora sojae
                     genes during infection and propagation_sHB"
                     /note="Vector: pBK-CMV; Site_1: EcoR1; Site_2: Xho1"
ORIGIN

  Query Match             8.1%;  Score 37;  DB 14;  Length 610;
  Best Local Similarity   51.5%;  Pred. No. 22;
  Matches   85;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy         26 TCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCG 85
              | | |||  | ||  ||  || | |||  || |  |  ||||      ||   | |||||
Db        212 TTGTAGCGGTAGTCCTTGATGTCCGTGAGCTTGGGGAAGTCGCCGATGAAGTAGCCGGCG 271

Qy         86 GCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCC 145
              || | || |  ||| ||||| | | |    | |       | ||||| .  | ||    |
Db        272 CCGTTGATGAGCACGACGGCGCCCACAAAGGCCGTGGCCGCGATGATGGCCAGGCGCTCG 331

Qy        146 GTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCC 190
              | |||   |||  |   ||| |||  |  ||  |||||  ||
Db        332 ATCTTGAAGGCCAGCAGGCCGCAGATGAGCCCCAGCACGATGGCC 376


RESULT 5
CNS02I89/c
LOCUS       CNS02I89                891 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            141G11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL198594
VERSION     AL198594.1  GI:7836745
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 891)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .891
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="141G11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG141AD06SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             8.0%;  Score 36.6;  DB 29;  Length 891;
  Best Local Similarity   51.2%;  Pred. No. 33;
  Matches   66;  Conservative    7;  Mismatches   56;  Indels    0;  Gaps    0;

Qy         51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
              ||  ||| | ||||| |    :  | |||  | ||||||| |       | || ||| :
Db        437 TGTCCTGCCGGGCGTAGGGGAWGCAGGGGATGACGGCGGTKACGCGTGACGAGGAGGCGW 378

Qy        111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
              |  | | |||   |: :|||:||||||| |    | :|||||     :|      |||
Db        377 TCTTGCAGGCGTTGWTYATGWTCAGCTGCTCCWTCAWGTTGTCGTTAWTTTCGCCGCAGC 318

Qy        171 TGCACTGCA 179
              || |||||
Db        317 CGCTCTGCA 309
```

RESULT 6
BZ568375
LOCUS BZ568375 787 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_7503.x1 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_7503, genomic survey sequence.
ACCESSION BZ568375
VERSION BZ568375.1 GI:27201285
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 787)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .787
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_7503"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."
ORIGIN

Query Match 8.0%; Score 36.4; DB 28; Length 787;
Best Local Similarity 47.4%; Pred. No. 36;
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```
Qy        17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
             | | || ||| |  |||| |   | |  || | || ||  ||| | | |
Db       146 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 205

Qy        77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
             |||||||| | |   ||  ||||| || | |    || || ||    |||
Db       206 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGCTCGACCAGGGTGGCGATGTTG 265

Qy       137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
             | |  |||    || | |||| | | |    | |||   ||||  |  | |
Db       266 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 325

Qy       197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
               ||     || ||  | |  |||| | |  | ||||   || ||||||
Db       326 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAAACCGGCCTTGACGTAGGC 375
```

RESULT 7
FR0020317
LOCUS FR0020317 417 bp DNA linear GSS 10-DEC-1997
DEFINITION F.rubripes GSS sequence, clone 040D20bD10, genomic survey sequence.
ACCESSION AL013202
VERSION AL013202.1 GI:2679570
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 417)
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y., Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic sequence.
FEATURES Location/Qualifiers
source 1. .417
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="040D20bD10"
/clone_lib="cosmid 040D20"
ORIGIN

Query Match 7.7%; Score 35.2; DB 29; Length 417;
Best Local Similarity 50.3%; Pred. No. 59;
Matches 76; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy       170 ATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
             |||||   ||||  |  || |   | || |  |||  |   || ||   |||||  ||| |
Db        16 ATGCAANCCAAAAAGGAGCACCNCANTCCGTCACGCGCNCAACGCACGCCGGTTCAGGGA 75

Qy       230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
             |  |||  |||   |  |  || |      | |   | | |  |||||  |  |  |
Db        76 ATTCNGCCTGTGTTCACACGCCTGTGTTCTTCATTTGGACTCCAAGGACTCTAATACNTG 135

Qy       290 TCAGAAATAATGCCACCATCGACCAGTGGAA 320
               |||||  ||   ||||  ||||   ||||
Db       136 AGAGAAACGATTATACCACGGACCTNAGGAA 166
```

RESULT 8
BZ469000/c
LOCUS BZ469000 811 bp DNA linear GSS 13-DEC-2002
DEFINITION BOOAO58TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOAO58, genomic survey sequence.
ACCESSION BZ469000
VERSION BZ469000.1 GI:26764546
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 811)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOOAO58TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .811
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOOAO58"
/clone_lib="BO_1.6_2_KB_tot"

/note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 811;
Best Local Similarity 47.7%; Pred. No. 78;
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```
Qy        146 GTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGA 205
              || |||  |   | ||    | |  |    ||  | || |     ||||| |||
Db        640 GTATTGAAAATAAGGATTGTGAAGCTTGCATGTTAGGCAAGCATTGTAAAACTGTGTTTC 581

Qy        206 CCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTA 265
              |||||  |    |  | |||||  ||  ||   ||| ||   ||   ||  | |   | ||
Db        580 CCATTTCATCTACTGTTTATGACAAATGCTTTGATCTAATTCACTCTGATGTTTGGACTG 521

Qy        266 CTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
              ||  |    || ||  | || |||| | | |   || |||  ||    |    ||  | |
Db        520 CTCCTTGCTTGTCTAGGGATAGTTTTAAATACTTTGTCACATTCATAAATGAAAAATCGA 461

Qy        326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
              || || ||   || ||  |    |||  |   | ||
Db        460 AATACACCTGGATAACACTTATTCAAACCAAAGATA 425
```

RESULT 9
BZ204853
LOCUS BZ204853 813 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-390D21, genomic survey sequence.
ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 813)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-390D21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 390 row: D column: 21
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .813
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"
ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 813;
Best Local Similarity 49.0%; Pred. No. 78;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```
Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
              ||| |   | | | | || |    |  ||   ||  | ||  |  |   ||  |  ||
Db        548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              || |   || ||       || |  | |  |||||  |     |||||  ||  |   |
Db        608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              ||   ||  |||||||| ||| |    | ||||| | |   | |||| ||| ||      |
Db        668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAAGGACAC 727

Qy        358 GGTAATAACGCC 369
              ||  | | | |
Db        728 TGTGGTTAGGAC 739
```

RESULT 10
CA588919/c
LOCUS CA588919 490 bp mRNA linear EST 19-NOV-2002
DEFINITION hab51c12.y1 Fugu UT7 adult skin Takifugu rubripes cDNA clone IMAGE:6352126 5' similar to SW:CLD4_MOUSE O35054 CLAUDIN-4 ;, mRNA sequence.
ACCESSION CA588919
VERSION CA588919.1 GI:25133497
KEYWORDS EST.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 490)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
Library materials provided by G. Elgar (UK MRC HGMP-RC) Library constructed by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi and S. Watabe (University of Tokyo, Institute of Medical Science and Laboratory of Aquatic Molecular Biology and Biotechnology) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Fugu clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham.
FEATURES Location/Qualifiers
source 1. .490
/organism="Takifugu rubripes"
/mol_type="mRNA"
/db_xref="taxon:31033"
/clone="IMAGE:6352126"
/sex="female"
/tissue_type="skin"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="Fugu UT7 adult skin"

/note="Vector: pBluescript-FL; Site_1: PflM I (CCANNNNNTGG); Site_2: PflM I (CCANNNNNTGG); BamHI-SmaI sites were converted to BamHI-PflMI-SfiI-PflMI sites (SmaI is destroyed). Other part of the vector is untouched. The cDNA is inserted between two PflMI sites in T3 (5') to T7 (3') direction. Library materials provided by G. Elgar (UK MRC HGMP-RC) and constructed and donated by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi and S. Watabe (University of Tokyo, Institute of Medical Science and Laboratory of Aquatic Molecular Biology and Biotechnology)."

ORIGIN

Query Match 7.7%; Score 35; DB 14; Length 490;
Best Local Similarity 51.6%; Pred. No. 72;
Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        92 ATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTG 151
             ||||| | ||||  ||||      | |||||   | |  |||| ||    |
Db       307 ATCATGAGGGCGCGGGCAGCCTGGAGGTCCCGGGAAAGGGCCAGCAGGGAGTCGTAAACC 248

Qy       152 TTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTA 211
             || | |  |  || |  ||| |||||   ||  |  | | |  |       | | |
Db       247 TTGCACTGCATCTGACCTGTGCTCTGCACCACGCAGTTCATCCAGATCCCTTCCCAAGTG 188

Qy       212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
                 |||| || ||| |||||| ||||| | ||||
Db       187 ATTTGAGCCGTGATGATAACGGGGCCGACGAAGGC 153
```

RESULT 11
AU085223
LOCUS AU085223 480 bp mRNA linear EST 01-SEP-2000
DEFINITION AU085223 Cryptomeria japonica inner bark Cryptomeria japonica cDNA clone CC2407 5', mRNA sequence.
ACCESSION AU085223
VERSION AU085223.1 GI:9963206
KEYWORDS EST.
SOURCE Cryptomeria japonica (Japanese cedar)
ORGANISM Cryptomeria japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.
REFERENCE 1 (bases 1 to 480)
AUTHORS Ujino-Ihara,T., Yoshimura,K., Ugawa,Y., Yoshimaru,H., Nagasaka,K. and Tsumura,Y.
TITLE Expression analysis of ESTs derived from the inner bark of Cryptomeria japonica
JOURNAL Plant Mol. Biol. 43 (4), 451-457 (2000)
MEDLINE 20505378
PUBMED 11052197
COMMENT Contact: Tokuko Ujino-Ihara
Bio-resources Technology Division
Forestry and Forest Products Research Institute
Matsunosato 1, Kukizaki, Ibaraki 305-8687, Japan
Tel: 81-298-73-3211(ex.444)
Fax: 81-298-73-3795
Email: udino@ffpri.affrc.go.jp;
URL:http://www.ffpri.affrc.go.jp/labs/cjgenome/database/cjdatae.htm
1.
FEATURES Location/Qualifiers
source 1. .480
/organism="Cryptomeria japonica"
/mol_type="mRNA"
/db_xref="taxon:3369"
/clone="CC2407"
/tissue_type="inner bark"
/clone_lib="Cryptomeria japonica inner bark"

ORIGIN

Query Match 7.6%; Score 34.8; DB 9; Length 480;
Best Local Similarity 51.3%; Pred. No. 81;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy       147 TGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGAC 206
             || | ||| | ||  ||| | ||    |   ||   || ||  |   |||| |||||  |||
Db       219 TGGTTTTGCTCAACAAAGGCCAGGGGTTATGGGTAGAGAATCAAATAAAACTGAAGGGAC 278

Qy       207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
                 || ||||| ||| |   | | ||        ||| ||  |   |||    ||||  ||
Db       279 TGCGACACAGAGAGGTCACTTTGAGGGGCATTCTGTTGGAAATTCTGCCCGGAATATGAC 338

Qy       267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCA 304
             |  |       | || ||   |  || | | ||| ||
Db       339 TGGTAGCAATTCACAAAAGTATCAAAGCATTCATGGCA 376
```

RESULT 12
AU241511/c
LOCUS AU241511 406 bp mRNA linear EST 15-JAN-2002
DEFINITION AU241511 UV irradiated OLHNI cell line cDNA library (OLc) Oryzias latipes cDNA clone OLc53.01d similar to pir|T16763|hypothetical protein R151.2 - Caenorhabditis elegans, mRNA sequence.
ACCESSION AU241511
VERSION AU241511.1 GI:18154090
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 406)
AUTHORS Naruse,K., Mitani,H. and Tanaka,M.
TITLE Medaka EST Project in University of Tokyo (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from UV irradiated OLHNI cell line cDNA library (OLc) 5' end sequences.
FEATURES Location/Qualifiers
source 1. .406
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="HNI"
/db_xref="taxon:8090"
/clone="OLc53.01d"
/clone_lib="UV irradiated OLHNI cell line cDNA library (OLc)"

ORIGIN

Query Match 7.6%; Score 34.6; DB 9; Length 406;
Best Local Similarity 54.3%; Pred. No. 86;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy        51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
             ||  ||| | |||||       | | |||  | ||||||| | |       | | ||| |
Db       293 TGTCCTGCCGGGCGTACGGGAAGCAGGGGATGACGGCGGTGACCCGGGACGACGAGGCGA 234

Qy       111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
             |  | | |||   ||  |||||||||| | |    |  |||||    ||   |  |||
Db       233 TCTTGCAGGCGTTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC 174

Qy       171 TGCACTGCA 179
                 || |||||
Db       173 CGCTCTGCA 165
```

RESULT 13
BJ006097/c
LOCUS BJ006097 499 bp mRNA linear EST 05-DEC-2001

```
DEFINITION  BJ006097 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA088D11 5',
            mRNA sequence.
ACCESSION   BJ006097
VERSION     BJ006097.1  GI:17359814
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA088D11"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 499;
  Best Local Similarity   54.3%;  Pred. No. 94;
  Matches   70;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy         51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
              ||  ||| | |||||       |  | |||  | |||||||| | |       |  | ||| |
Db        195 TGTCCTGCCGGGCGTACGGGAAGCAGGGGATGACGGCGGTGACCCGGGACGACGAGGCGA 136

Qy        111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
              |  | | |||   ||  |||||||||| | |     |  |||||      ||   |  |||
Db        135 TCTTGCAGGCGTTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC 76

Qy        171 TGCACTGCA 179
                || |||||
Db         75 CGCTCTGCA 67


RESULT 14
BJ001420/c
LOCUS       BJ001420                561 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ001420 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA007C02 5',
            mRNA sequence.
ACCESSION   BJ001420
VERSION     BJ001420.1  GI:17361687
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA007C02"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 561;
  Best Local Similarity   54.3%;  Pred. No. 98;
  Matches   70;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy         51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
              ||  ||| | |||||       |  | |||  | |||||||| | |       |  | ||| |
Db        230 TGTCCTGCCGGGCGTACGGGAAGCAGGGGATGACGGCGGTGACCCGGGACGACGAGGCGA 171

Qy        111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
              |  | | |||   ||  |||||||||| | |     |  |||||      ||   |  |||
Db        170 TCTTGCAGGCGTTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC 111

Qy        171 TGCACTGCA 179
                || |||||
Db        110 CGCTCTGCA 102


RESULT 15
BJ000263/c
LOCUS       BJ000263                640 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ000263 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA014A01 5',
            mRNA sequence.
ACCESSION   BJ000263
VERSION     BJ000263.1  GI:17358166
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA014A01"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 640;
```

Best Local Similarity 54.3%; Pred. No. 1e+02;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
               ||  ||| | |||||      |  | |||  | |||||||| | |       |  | ||| |
Db         387 TGTCCTGCCGGGCGTACGGGAAGCAGGGGATGACGGCGGTGACCCGGGACGACGAGGCGA 328

Qy         111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
               |  | | |||   ||  |||||||||| | |     | |||||      ||   |  |||
Db         327 TCTTGCAGGCGTTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC 268

Qy         171 TGCACTGCA 179
                || |||||
Db         267 CGCTCTGCA 259
```

Search completed: March 18, 2004, 20:11:31
Job time : 2036.55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		412.8	90.5	456	6	I44909	I44909 Sequence 58
2		412.8	90.5	2067	1	SEU43280	U43280 Salmonella
3		404.8	88.8	5103	1	STAJ2301	AJ002301 Salmonell
4		404.8	88.8	22411	1	AE008749	AE008749 Salmonell
5		403.2	88.4	254050	1	AL627269	AL627269 Salmonell
6	c	403.2	88.4	301983	1	AE016840	AE016840 Salmonell
7		397.4	87.1	1048	1	STAGFBA	AJ000514 Salmonell
8		288.6	63.3	361	6	I44908	I44908 Sequence 56
9		280	61.4	2889	1	CSP515700	AJ515700 Citrobact
10		238.4	52.3	4680	1	ECCSGABDG	X90754 E.coli csgG
11		238.4	52.3	10346	1	AE000205	AE000205 Escherich
12		238.4	52.3	15047	1	D90741	D90741 Escherichia
13		237.2	52.0	2920	1	CFR515701	AJ515701 Citrobact
14		236.8	51.9	456	6	AX814811	AX814811 Sequence
15		236.8	51.9	648	1	ECOCSGAA	L04979 Escherichia
16		228.6	50.1	1711	1	AF275733	AF275733 Escherich
17		228.6	50.1	10190	1	AE005315	AE005315 Escherich
18		228.6	50.1	327773	1	AP002554	AP002554 Escherich
19		227	49.8	306358	1	AE016759	AE016759 Escherich
20		195.2	42.8	2883	1	ESA515702	AJ515702 Enterobac
21		185.8	40.7	230	1	SEU53207	U53207 Salmonella
22		169.6	37.2	10370	1	AE015131	AE015131 Shigella
23		169.6	37.2	292504	1	AE016981	AE016981 Shigella
24	c	146.6	32.1	437	1	AF237726	AF237726 Shigella
25		93.2	20.4	19201	1	D90742	D90742 Escherichia
26		64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27		48.2	10.6	78	6	AX814809	AX814809 Sequence
28		44.8	9.8	72	6	AX814798	AX814798 Sequence
29	c	42	9.2	2000	6	AX655393	AX655393 Sequence
30	c	36.2	7.9	110243	2	AC146945	AC146945 Oryza sat
31	c	36.2	7.9	148860	8	OSJN00093	AL606648 Oryza sat
32		35.8	7.9	298900	1	AP005937	AP005937 Bradyrhiz
33	c	35.6	7.8	153232	9	AC024086	AC024086 Homo sapi
34	c	35.6	7.8	185124	9	AC009358	AC009358 Human Chr
35	c	35.6	7.8	186208	9	AC142338	AC142338 Pan trogl
36	c	35.4	7.8	2916	14	FAD459805	AJ459805 Fowl aden
37	c	35.4	7.8	2929	14	FAD554049	AJ554049 Fowl aden
38	c	35.4	7.8	3300	14	FAD431719	AJ431719 Fowl aden
39	c	35.4	7.8	3315	14	FAU26221	U26221 Fowl adenov
40	c	35	7.7	5014	3	PFPFPRNA	X63648 P.falciparu
41	c	35	7.7	12641	1	AE014731	AE014731 Bifidobac
42	c	35	7.7	349980	6	AX492785	AX492785 Sequence
43	c	35	7.7	349980	6	AX553952	AX553952 Sequence
44	c	34.8	7.6	301727	1	AE016960	AE016960 Coxiella
45		34.8	7.6	347625	1	BX248356	BX248356 Corynebac

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match              90.5%;  Score 412.8;  DB 6;  Length 456;
  Best Local Similarity    94.1%;  Pred. No. 6.6e-108;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||   | |    | |   ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match              90.5%;  Score 412.8;  DB 1;  Length 2067;
  Best Local Similarity    94.1%;  Pred. No. 7.2e-108;
```

```
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |     | |    ||  ||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match          88.8%;  Score 404.8;  DB 1;  Length 5103;
  Best Local Similarity  93.0%;  Pred. No. 1.6e-105;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  |  |||||  | |  | |  || ||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA

COMMENT COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES Location/Qualifiers

source 1. .22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

gene 434. .1308
/gene="phoH"
/note="synonym: STM1126"

RBS 434. .439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"

CDS 454. .1308
/gene="phoH"
/note="similar to E. coli PhoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=11
/product="PhoB-dependent ATP-binding pho regulon component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
QRTLHAYS"

gene complement(1414. .2303)
/gene="STM1127"

CDS complement(1414. .2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29% identity in aa 15 - 264"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
QARESKLRTMNALDMTK"

RBS complement(2298. .2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"

gene complement(2580. .4088)
/gene="STM1128"

CDS complement(2580. .4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22% identity in aa 7 - 478"
/codon_start=1
/transl_table=11
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"

RBS complement(4083. .4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"

gene complement(4413. .5099)
/gene="STM1129"

CDS complement(4413. .5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 - 226"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
WFSHAVKR"

RBS complement(5094. .5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"

gene 5599. .6759
/gene="STM1130"

CDS 5599. .6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41% identity in aa 40 - 403"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"

gene 6793. .7497
/gene="STM1131"

RBS 6793. .6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match             88.8%;  Score 404.8;  DB 1;  Length 22411;
  Best Local Similarity   93.0%;  Pred. No. 1.7e-105;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |     || ||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             88.4%;  Score 403.2;  DB 1;  Length 254050;
  Best Local Similarity   92.8%;  Pred. No. 5.7e-105;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||| ||| | |||||| | | | || ||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840               301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRPEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match           88.4%;  Score 403.2;  DB 1;  Length 301983;
Best Local Similarity 92.8%;  Pred. No. 5.8e-105;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |     | |    ||  ||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match           87.1%;  Score 397.4;  DB 1;  Length 1048;
  Best Local Similarity   92.1%;  Pred. No. 1.9e-103;
  Matches  419;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  ||||||  | |  | |  || ||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  ||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           63.3%;  Score 288.6;  DB 6;  Length 361;
  Best Local Similarity   91.3%;  Pred. No. 4.5e-72;
  Matches  306;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              |||||||||||||| |||  |  ||||||  | |  | |  || |||  |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           61.4%;  Score 280;  DB 1;  Length 2889;
  Best Local Similarity  77.4%;  Pred. No. 1.6e-69;
  Matches  353;  Conservative    0;  Mismatches  100;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||    |      ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||    |  |||||| ||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||| || ||||||| |||| |  ||  |  ||  |  | |    ||| |
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || |||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || |||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
```

```
                /db_xref="SWISS-PROT:P52104"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene            complement(1434. .1823)
                /gene="csgE"
CDS             complement(1434. .1823)
                /gene="csgE"
                /codon_start=1
                /transl_table=11
                /product="assembly /transport component in curli
                production"
                /protein_id="CAA62279.1"
                /db_xref="GI:1147561"
                /db_xref="SWISS-PROT:P52105"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
gene            complement(1828. .2478)
                /gene="csgD"
CDS             complement(1828. .2478)
                /gene="csgD"
                /note="necessary for transcription of the csgBA operon"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="CAA62280.1"
                /db_xref="GI:1147562"
                /db_xref="GOA:P52106"
                /db_xref="SWISS-PROT:P52106"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
misc_feature    complement(2626)
                /note="transcriptional start site for csgDEFG operon"
misc_feature    3140
                /note="transcriptional start site for csgBA operon"
gene            3233. .3688
                /gene="csgB"
CDS             3233. .3688
                /gene="csgB"
                /codon_start=1
                /transl_table=11
                /product="nucleation component of curlin monomers"
                /protein_id="CAA62281.1"
                /db_xref="GI:1147563"
                /db_xref="GOA:P39828"
                /db_xref="SWISS-PROT:P39828"
                /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
gene            3729. .4184
                /gene="csgA"
CDS             3729. .4184
                /gene="csgA"
                /note="major subunit"
                /codon_start=1
                /transl_table=11
                /product="curlin/CsgA protein"
                /protein_id="CAA62282.1"
                /db_xref="GI:1147564"
                /db_xref="GOA:P28307"
                /db_xref="SWISS-PROT:P28307"
                /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
sig_peptide     3729. .3788
                /gene="csgA"
gene            4243. .4575
                /gene="orfC"
CDS             4243. .4575
                /gene="orfC"
                /codon_start=1
                /transl_table=11
                /product="putative curli production protein"
                /protein_id="CAA62283.1"
                /db_xref="GI:1147565"
                /db_xref="SWISS-PROT:P52107"
                /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                WPPSSEKS"
ORIGIN

  Query Match           52.3%;  Score 238.4;  DB 1;  Length 4680;
  Best Local Similarity  70.2%;  Pred. No. 1.6e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  |||||||||| | |||   |    ||||||    ||    |  | |   ||| ||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 11
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
```

```
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTPLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
```

```
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

  Query Match           52.3%;  Score 238.4;  DB 1;  Length 10346;
  Best Local Similarity  70.2%;  Pred. No. 1.7e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  |||||||||| | |||   |    ||||||    ||    |  | |    ||| ||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448


RESULT 12
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                    Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES                Location/Qualifiers
     source             1. .15047
                        /organism="Escherichia coli K12"
                        /mol_type="genomic DNA"
                        /strain="K12"
                        /db_xref="taxon:83333"
                        /map="23.7-24.0 min"
                        /clone="Kohara clone #231"
                        /note="Nucleotide position 1097101-1112147 from the
```

initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDPTIGILGAGVLGSKVAQSLQTWRF PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS 2906. .3460
/note="ORF_ID:o231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=11
/product="Hypothetical protein HI1543"
/protein_id="BAA35825.1"
/db_xref="GI:4062608"
/translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW RTMAPLTRDAISAMWDELEEDSEE"
CDS 3514. .4053
/note="unnamed protein product; ORF_ID:o231#4"
/codon_start=1
/transl_table=11
/protein_id="BAA35826.1"
/db_xref="GI:4062609"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW LAARSAKTAHREQEIKNKA"
gene complement(4117. .4950)
/gene="csgG"
CDS complement(4117. .4950)
/gene="csgG"
/note="ORF_ID:o231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component CsgG precursor."
/protein_id="BAA35827.1"
/db_xref="GI:4062610"
/translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE S"
gene complement(4977. .5393)
/gene="csgF"
CDS complement(4977. .5393)
/gene="csgF"
/note="ORF_ID:o231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component CsgF precursor."
/protein_id="BAA35828.1"
/db_xref="GI:4062611"
/translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene complement(5418. .5807)
/gene="csgE"
CDS complement(5418. .5807)
/gene="csgE"
/note="ORF_ID:o231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component CsgE precursor."
/protein_id="BAA35829.1"
/db_xref="GI:4062612"
/translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL IQTEEALNRRQINQALLSTGDLAHDEF"
gene complement(5812. .6462)
/gene="csgD"
CDS complement(5812. .6462)
/gene="csgD"
/note="ORF_ID:o231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=11
/product="Probable csgAB operon transcriptional regulatory protein."
/protein_id="BAA35830.1"
/db_xref="GI:4062613"
/translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene 7217. .7672
/gene="csgB"
CDS 7217. .7672

Query Match 52.3%; Score 238.4; DB 1; Length 15047;
Best Local Similarity 70.2%; Pred. No. 1.7e-57;
Matches 320; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db 7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy 61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db 7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
Db 7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy 181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
Db 7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

```
Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | |||||| |||||| | |||  ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168
```

RESULT 13
CFR515701
LOCUS CFR515701 2920 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION AJ515701
VERSION AJ515701.1 GI:31790495
KEYWORDS csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation component of curlin monomers; regulatory protein.
SOURCE Citrobacter freundii
ORGANISM Citrobacter freundii
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.
REFERENCE 1
AUTHORS Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
JOURNAL Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE 2 (bases 1 to 2920)
AUTHORS Romling,U.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES Location/Qualifiers
source 1. .2920
/organism="Citrobacter freundii"
/mol_type="genomic DNA"
/isolate="Fec4"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:546"
/country="Germany"
gene complement(211. .861)
/gene="csgD"
CDS complement(211. .861)
/gene="csgD"
/function="necessary for transcription of the csgBA operon"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56673.1"
/db_xref="GI:31790496"
/translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene 1624. .2082
/gene="csgB"
CDS 1624. .2082
/gene="csgB"
/function="nucleator"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56674.1"
/db_xref="GI:31790497"
/translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
gene 2123. .2572
/gene="csgA"
CDS 2123. .2572
/gene="csgA"
/function="curli subunit"
/codon_start=1
/transl_table=11
/product="curlin-csgA protein"
/protein_id="CAD56675.1"
/db_xref="GI:31790498"
/translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"

ORIGIN

Query Match 52.0%; Score 237.2; DB 1; Length 2920;
Best Local Similarity 71.7%; Pred. No. 3.5e-57;
Matches 327; Conservative 0; Mismatches 123; Indels 6; Gaps 1;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| ||||||||||||          |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||    ||| ||||   ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              || ||||| |||||||  |||         | || |     |          |   ||  |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |||||||||||| ||| | |||||  |||    ||||  ||||  || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||  ||||| ||||||||  || |||||   | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||| |||||||||||||| || ||||| ||||  ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

RESULT 14
AX814811
LOCUS AX814811 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
TITLE Peptides
JOURNAL Patent: WO 03064446-A 15 07-AUG-2003;
Hansa Medical Research Aktiebolag (SE)
FEATURES Location/Qualifiers
source 1. .456

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           51.9%;  Score 236.8;  DB 6;  Length 456;
  Best Local Similarity  70.0%;  Pred. No. 4.1e-57;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| | |||| | ||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | ||||||||||||| ||    || ||||| |||| | ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              | |||||||||| | |||   |    ||||||    ||     | | |    ||| ||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match           51.9%;  Score 236.8;  DB 1;  Length 648;
  Best Local Similarity  70.0%;  Pred. No. 4.1e-57;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | ||||||||||||| ||    || ||||| |||| | ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              | |||||||||| | |||   |    ||||||    ||     | | |    ||| ||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 18, 2004, 15:02:54
Job time : 2312.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64626	Aac64626 AgfA::PT3
	2	412.8	90.5	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	412.8	90.5	456	2	AAT74142	Aat74142 Salmonell
	4	412.8	90.5	456	3	AAC64617	Aac64617 Salmonell
	5	385.6	84.6	456	3	AAC64629	Aac64629 AgfA::PT3
	6	363.2	79.6	456	3	AAC64625	Aac64625 AgfA::PT3
	7	361.6	79.3	456	3	AAC64628	Aac64628 AgfA::PT3
	8	361.6	79.3	456	3	AAC64622	Aac64622 AgfA::PT3
	9	361.6	79.3	456	3	AAC64623	Aac64623 AgfA::PT3
	10	360	78.9	456	3	AAC64630	Aac64630 AgfA::PT3
	11	360	78.9	456	3	AAC64624	Aac64624 AgfA::PT3
	12	358.4	78.6	456	3	AAC64627	Aac64627 AgfA::PT3
	13	355.2	77.9	456	3	AAC64631	Aac64631 AgfA::PT3
	14	288.6	63.3	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	288.6	63.3	361	2	AAT74141	Aat74141 Salmonell
	16	236.8	51.9	456	3	AAC64619	Aac64619 Escherich
	17	236.8	51.9	456	9	ACF36153	Acf36153 E. coli C
	18	224.4	49.2	646	2	AAQ62647	Aaq62647 Fibronect
	19	162.4	35.6	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
c	21	50	11.0	78	3	AAC64609	Aac64609 AgfA (SEF
	22	50	11.0	78	3	AAC64610	Aac64610 AgfA (SEF
	23	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	24	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	25	45	9.9	78	3	AAC64606	Aac64606 SefA (SEF
	26	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
c	27	43.4	9.5	78	3	AAC64605	Aac64605 SefA (SEF
	28	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	29	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
c	30	42	9.2	2000	7	ADA71938	Ada71938 Rice gene
c	31	35	7.7	349980	6	ABQ81847	Abq81847 Bifidobac
c	32	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	33	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	34	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	35	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	36	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	37	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c	38	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a
c	39	33	7.2	110000	6	ABA90521_17	Continuation (18 o
	40	32.8	7.2	3562	4	ABL06749	Abl06749 Drosophil
	41	32.8	7.2	24081	7	AAD54223	Aad54223 Streptomy
	42	32.8	7.2	52101	7	AAD54217	Aad54217 Streptomy
	43	32.6	7.1	100	7	ACD68809	Acd68809 E. coli K
	44	32.6	7.1	2553	3	AAA09500	Aaa09500 V. furnis
c	45	32.4	7.1	901	4	ABL09481	Abl09481 Drosophil

ALIGNMENTS

RESULT 1
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.2e-139;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 2; Length 456;
Best Local Similarity 94.1%; Pred. No. 3.6e-125;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| | |||||| | | | | || ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 412.8;  DB 2;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 3.6e-125;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

```
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 412.8;  DB 3;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 3.6e-125;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 5
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             84.6%;  Score 385.6;  DB 3;  Length 456;
  Best Local Similarity   90.4%;  Pred. No. 3.1e-116;
  Matches  412;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |   ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
                | | |||||      |   ||      |  | | ||  | |      | |    ||  ||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 6
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match            79.6%;  Score 363.2;  DB 3;  Length 456;
  Best Local Similarity  87.3%;  Pred. No. 7.3e-109;
  Matches  398;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |    ||   |      | | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 79.3%; Score 361.6; DB 3; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.5e-108;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  ||||||  | |  | |  || ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||  ||  ||  |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |  || |  |||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 8
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 79.3%; Score 361.6; DB 3; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.5e-108;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||| |          | | || | ||  |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 9
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match   79.3%; Score 361.6; DB 3; Length 456;
  Best Local Similarity   87.1%; Pred. No. 2.5e-108;
  Matches 397; Conservative   0; Mismatches   59; Indels   0; Gaps   0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||     | ||||  ||         |  || | ||||||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 78.9%; Score 360; DB 3; Length 456;
Best Local Similarity 86.8%; Pred. No. 8.3e-108;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  |||||| | |  | |  || ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | |||  |    | |  ||  |   |    |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

  Query Match             78.9%;  Score 360;  DB 3;  Length 456;
  Best Local Similarity   86.8%;  Pred. No. 8.3e-108;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||   | | | |   | |   |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||   | |   | |   ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match             78.6%;  Score 358.4;  DB 3;  Length 456;
  Best Local Similarity   86.6%;  Pred. No. 2.8e-107;
  Matches  395;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||   | |   | |   ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||  |   |         | ||  |    ||  |   | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 3.1e-106;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||   | |     | |   ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  || |||  |   |          |   | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             63.3%;  Score 288.6;  DB 2;  Length 361;
  Best Local Similarity   91.3%;  Pred. No. 2.3e-84;
  Matches  306;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              |||||||||||| ||| |  |  ||||||  | |  | |  || ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag= b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag= c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag= d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag= e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             63.3%;  Score 288.6;  DB 2;  Length 361;
  Best Local Similarity   91.3%;  Pred. No. 2.3e-84;
  Matches  306;  Conservative    0;  Mismatches   29;  .Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              |||||||||||| ||| |  |  ||||||  | |  | |  || ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 18, 2004, 09:12:00
```

Job time : 289.405 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	412.8	90.5	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	288.6	63.3	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	32.8	7.2	1218	4	US-09-489-039A-403	Sequence 403, App
c	6	32.2	7.1	399	4	US-09-621-976-8976	Sequence 8976, Ap
	7	32	7.0	1152	4	US-09-489-039A-2755	Sequence 2755, Ap
c	8	31.4	6.9	534	4	US-09-489-039A-2195	Sequence 2195, Ap
	9	31.4	6.9	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
c	10	31.2	6.8	549	4	US-09-252-991A-14225	Sequence 14225, A
c	11	31.2	6.8	696	4	US-09-252-991A-14137	Sequence 14137, A
c	12	31	6.8	2052	1	US-08-443-104-5	Sequence 5, Appli
c	13	31	6.8	2052	1	US-08-238-130-6	Sequence 6, Appli
c	14	31	6.8	2052	1	US-08-442-859-5	Sequence 5, Appli
c	15	31	6.8	2052	2	US-08-398-489-5	Sequence 5, Appli
c	16	31	6.8	2052	2	US-08-894-772-1	Sequence 1, Appli
c	17	31	6.8	2052	2	US-09-207-844-1	Sequence 1, Appli
c	18	31	6.8	2052	4	US-09-252-509-1	Sequence 1, Appli
c	19	31	6.8	2052	5	PCT-US95-05534-5	Sequence 5, Appli
	20	30.8	6.8	1662	1	US-08-565-386-2	Sequence 2, Appli
	21	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
	22	30.8	6.8	4529	1	US-08-565-386-1	Sequence 1, Appli
	23	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	24	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	25	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	26	30	6.6	2436	4	US-09-540-236-492	Sequence 492, App
c	27	30	6.6	3895	4	US-08-961-527-201	Sequence 201, App
c	28	30	6.6	65792	4	US-09-596-002-31	Sequence 31, Appl
	29	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	30	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	31	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
	32	29.8	6.5	51259	3	US-08-781-891-209	Sequence 209, App
	33	29.8	6.5	51259	4	US-09-618-166-209	Sequence 209, App
c	34	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
	35	29.4	6.4	912	4	US-09-543-681A-2040	Sequence 2040, Ap
c	36	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	37	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	38	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	39	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
	40	29.2	6.4	708	4	US-09-543-681A-2551	Sequence 2551, Ap
	41	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	42	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	43	29	6.4	939	4	US-09-976-594-108	Sequence 108, App
c	44	29	6.4	1097	4	US-09-988-784-7	Sequence 7, Appli
c	45	29	6.4	98844	4	US-09-791-211-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 90.5%; Score 412.8; DB 1; Length 456;
Best Local Similarity 94.1%; Pred. No. 1.7e-137;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||| ||| |  |||||| | |  | |  || ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             63.3%;  Score 288.6;  DB 1;  Length 361;
  Best Local Similarity   91.3%;  Pred. No. 3.8e-93;
  Matches  306;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              |||||||||||||| ||| |  |||||| | |  | |  || ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

```
RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
;  GENERAL INFORMATION:
;    APPLICANT:  Kim, Jihyun Francis
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;    TITLE OF INVENTION:  ERWINIA AMYLOVORA AND ITS USE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,927
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,108
;      FILING DATE:  06-AUG-1977
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1581
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
```

```
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1344 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.09;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||.| |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.09;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||.| |  ||   |    |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-489-039A-403
; Sequence 403, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 403
;   LENGTH: 1218
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-403

  Query Match             7.2%;  Score 32.8;  DB 4;  Length 1218;
  Best Local Similarity   58.0%;  Pred. No. 0.23;
  Matches   58;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        174 TCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAAT 233
              ||  || |||| ||  | | ||||||| |||||   ||   |||| |||      || ||
Db        614 TCGTCACAGCGCTGATCCTCAATATGTTCAGCACCTTTATCGTGCTGTCGGTAATTAACC 673

Qy        234 GGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
              |      || || ||| | || |  ||||  |||   |||
Db        674 CGACGCGGCCGGGCAGCGAGCAGGAAATAAAACTCGAGAA 713


RESULT 6
US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 8976
;   LENGTH: 399
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-8976

  Query Match             7.1%;  Score 32.2;  DB 4;  Length 399;
  Best Local Similarity   10.0%;  Pred. No. 0.19;
  Matches   34;  Conservative  151;  Mismatches  154;  Indels    0;  Gaps    0;

Qy         88 GGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 147
              | : ::::    |:::::: :|:  : :: ::   ::::::   |::| :      : :|
Db        340 GWKYWYYWMTSRGSYRRYKTSAMMGRAKMKKRKKTKKMYWMKGGGKKGGSTYMAMRSRRG 281

Qy        148 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTG 207
                 :  :   ::: ::::| |:::::  ::: ::::     :  :: :  ::|: ::|: :
Db        280 STGRWSYRRAMWRGSKSWGGGSYYRMAGYRSSRWRSWYSAMWRKKKMTCWKGRSSWGSRS 221

Qy        208 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACT 267
              :   : :: : | :  : :: :::  :  |:  : |::|: ::| :|:  : :   |
Db        220 TGYYAWMYKKSWCTSRKWMYYKKRRKKWRRKCTSTKRTCYRGSTYKCWKAYYTKKRRKWT 161

Qy        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327
              ::|   ::: : : : :  :| :::  ::::       |  |  |   :    : :  ::
Db        160 RWTYYYYKSYMSMKKTWRMKTAYYWTKRWKMTRTKWTWCTMCWKCTTYWMAGTMMYRYRR 101

Qy        328 AACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACC 387
             :::  :|:: ::||     :|: :: ::| :::: : :   | ::: ::|  ::  ::
Db        100 YWYYAKRAKWSKRCTWSTTCYCMKYMAKKCWSYWWSMSMMKWGKSMWWKWTYYYYYYMMK 41

Qy        388 GCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTT 426
              :   :   :|  : ::| : : ::   |:|: |::
Db         40 WSKMTYWSMMSCYARKCWRTYAKTYTWMTCMTWGKRTYK 2
```

RESULT 7
US-09-489-039A-2755
; Sequence 2755, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2755
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2755

Query Match 7.0%; Score 32; DB 4; Length 1152;
Best Local Similarity 56.7%; Pred. No. 0.43;
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCC 214
Db 656 CCACGGCCTCGGTGCTGGCTCTGCTCAAGCGGCTGCGTGATGAGTACCAGCTGGCGATCG 715

Qy 215 GTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGAT 258
Db 716 TCCTGATTACCCATGAAATGGACGCGGTGCGCACGGCGGCGGAT 759

RESULT 8
US-09-489-039A-2195/c
; Sequence 2195, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2195
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2195

Query Match 6.9%; Score 31.4; DB 4; Length 534;
Best Local Similarity 52.7%; Pred. No. 0.44;
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGT 133
Db 501 AAAGAGGCGGCGACGCAGCCGGTAGCGGTGCCTGCTGCACCGTCGGAAGCGACTCAGGGC 442

Qy 134 TGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTA 193
Db 441 GGCTTCTTTAGCCGCTGGCTTCAGGCGCTGTGGACGCTGTTCTGGAACACGATCCCGGTT 382

Qy 194 AATATGATC 202
Db 381 TATATCCTC 373

RESULT 9
US-09-489-039A-2162
; Sequence 2162, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2162
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (702)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-489-039A-2162

Query Match 6.9%; Score 31.4; DB 4; Length 1818;
Best Local Similarity 52.7%; Pred. No. 0.92;
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGT 133
Db 1363 AAAGAGGCGGCGACGCAGCCGGTAGCGGTGCCTGCTGCACCGTCGGAAGCGACTCAGGGC 1422

Qy 134 TGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTA 193
Db 1423 GGCTTCTTTAGCCGCTGGCTTCAGGCGCTGTGGACGCTGTTCTGGAACACGATCCCGGTT 1482

Qy 194 AATATGATC 202
Db 1483 TATATCCTC 1491

RESULT 10
US-09-252-991A-14225/c
; Sequence 14225, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14225
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14225

Query Match 6.8%; Score 31.2; DB 4; Length 549;
Best Local Similarity 53.2%; Pred. No. 0.53;
Matches 66; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
Db 282 CGTTGCACCGTCAGGCTCGGGGTCAGGCCGGTGCTGCACTGGATCTGGATATTGCCGCTG 223

Qy 191 GTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGG 250
Db 222 GTGCCGACGGTCTGCGCGTCGACCACGTTGGTAAGGTCGGAATGGCTGCCGAAGTCCAGG 163

```
Qy        251 GTGC 254
              || |
Db        162 GTTC 159


RESULT 11
US-09-252-991A-14137/c
; Sequence 14137, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 14137
;   LENGTH: 696
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14137

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 696;
  Best Local Similarity   53.2%;  Pred. No. 0.61;
  Matches   66;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
              |||||  |  | |       |||   ||    ||||| || | |||| | |  |  ||
Db        377 CGTTGCACCGTCAGGCTCGGGGTCAGGCCGGTGCTGCACTGGATCTGGATATTGCCGCTG 318

Qy        191 GTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGG 250
              ||    | | || ||  ||   ||  |||| ||      | |||||    | || |||||
Db        317 GTGCCGACGGTCTGCGCGTCGACCACGTTGGTAAGGTCGGAATGGCTGCCGAAGTCCAGG 258

Qy        251 GTGC 254
              || |
Db        257 GTTC 254


RESULT 12
US-08-443-104-5/c
; Sequence 5, Application US/08443104
; Patent No. 5691162
;  GENERAL INFORMATION:
;    APPLICANT:  Shuster, Jeffrey R.
;    APPLICANT:  Madden, Mark
;    APPLICANT:  Moyer, Donna L.
;    APPLICANT:  Fuglsang, Claus
;    APPLICANT:  Branner, Sven
;    TITLE OF INVENTION:  NOVEL METALLOPROTEASE HAVING INCREASED
;    TITLE OF INVENTION:  ACTIVITY
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5691162o No. 5691162disk of No. 5691162th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Tape
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/443,104
;      FILING DATE:  17-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/398,489
;      FILING DATE:  03=MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4180.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2052 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-443-104-5

  Query Match             6.8%;  Score 31;  DB 1;  Length 2052;
  Best Local Similarity   45.6%;  Pred. No. 1.4;
  Matches  109;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy        215 GTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              |||   ||| ||  |   | ||  | |    ||  |||  |   | |  ||     | |  |
Db        648 GTGAGCTTAGCCTCGGGCTTGCTGACGGTACCCTTGGTACCAGTGAAGGTGTAGTGGTTC 589

Qy        275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
              |  |||   |   |      || || | ||  ||  |  ||     || | |  ||
Db        588 TTGCTCTTCTTCTCAGCCTTGGCCTTGTCAGCCTCAACGGGGAGAGAAAGAACATCAACG 529

Qy        335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTG 394
              | |   ||   ||    |||| |    |  || | |||| | |||   |||||    |
Db        528 GTGTCCTTGAGAGCGTCGACGGGGTCTTTCTCGTCACGCTTGGTAAGAGGACCGGGAATC 469

Qy        395 ATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGTAT 453
              | ||  ||||   |    |  || ||    | |  |  || ||  || || | ||||
Db        468 TTGCCCTCGTAGAAGCTGTTTCCGTAGGAGAAGACCTCGCCGTCAGCGCCAATCTGTAT 410


RESULT 13
US-08-238-130-6/c
; Sequence 6, Application US/08238130
; Patent No. 5702934
;  GENERAL INFORMATION:
;    APPLICANT:  Hastrup, Sven
;    APPLICANT:  Branner, Sven
;    APPLICANT:  Jorgensen, Birthe R.
;    APPLICANT:  Christensen, Tove
;    APPLICANT:  Jorgensen, Birgitte B.
;    APPLICANT:  Shuster, Jeffrey R.
;    APPLICANT:  Madden, Mark
;    APPLICANT:  Moyer, Donna L.
;    APPLICANT:  Fuglsang, Claus
;    TITLE OF INVENTION:  PROCESSES FOR PRODUCING AN ENZYME
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5702934o No. 5702934disk of No. 5702934th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Tape
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/238,130
;      FILING DATE:  04-MAY-1994
;      CLASSIFICATION:  435
```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 522/93
; FILING DATE: 05-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3965.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1..363, 416..802, 856..1821, 1870..2052)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: join(1..363, 416..802, 856..1821, 1870..2049)
US-08-238-130-6

Query Match 6.8%; Score 31; DB 1; Length 2052;
Best Local Similarity 45.6%; Pred. No. 1.4;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;



RESULT 14
US-08-442-859-5/c
; Sequence 5, Application US/08442859
; Patent No. 5807729
; GENERAL INFORMATION:
; APPLICANT: Shuster, Jeffrey R.
; APPLICANT: Madden, Mark
; APPLICANT: Moyer, Donna L.
; APPLICANT: Fuglsang, Claus
; APPLICANT: Branner, Sven
; TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5807729o No. 5807729disk of No. 5807729th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,859
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 252
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,489
; FILING DATE: 03=MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 4180.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-442-859-5

Query Match 6.8%; Score 31; DB 1; Length 2052;
Best Local Similarity 45.6%; Pred. No. 1.4;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;



RESULT 15
US-08-398-489-5/c
; Sequence 5, Application US/08398489
; Patent No. 5843753
; GENERAL INFORMATION:
; APPLICANT: Shuster, Jeffrey R.
; APPLICANT: Madden, Mark
; APPLICANT: Moyer, Donna L.
; APPLICANT: Fuglsang, Claus
; APPLICANT: Branner, Sven
; TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5843753o No. 5843753disk of No. 5843753th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/398,489
; FILING DATE: 03-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris Dr., Cheryl H.

; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 4180.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-398-489-5

Query Match 6.8%; Score 31; DB 2; Length 2052;
Best Local Similarity 45.6%; Pred. No. 1.4;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

```
Qy        215 GTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              |||   ||| ||  |   | ||  | |    ||  ||| |   | |  ||    | | |
Db        648 GTGAGCTTAGCCTCGGGCTTGCTGACGGTACCCTTGGTACCAGTGAAGGTGTAGTGGTTC 589

Qy        275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
              |  |||    |    |        || || | || ||  |  ||    || | |  ||
Db        588 TTGCTCTTCTTCTCAGCCTTGGCCTTGTCAGCCTCAACGGGGAGAGAAAGAACATCAACG 529

Qy        335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTG 394
              | |   ||   ||    |||| |    |  || | |||| | |||   |||||    |
Db        528 GTGTCCTTGAGAGCGTCGACGGGGTCTTTCTCGTCACGCTTGGTAAGAGGACCGGGAATC 469

Qy        395 ATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGTAT 453
              | ||  ||||   |     |   || ||    | |  |  || ||  || || | ||||
Db        468 TTGCCCTCGTAGAAGCTGTTTCCGTAGGAGAAGACCTCGCCGTCAGCGCCAATCTGTAT 410
```

Search completed: March 18, 2004, 20:20:58
Job time : 62.4971 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	34.6	7.6	1662	14	US-10-156-761-3483	Sequence 3483, Ap
	2	34.6	7.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	3	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c	4	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c	5	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c	6	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c	7	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
c	8	33.4	7.3	655	14	US-10-184-644-418	Sequence 418, App
c	9	33.4	7.3	655	14	US-10-184-634-418	Sequence 418, App
c	10	33.2	7.3	1848	14	US-10-156-761-7222	Sequence 7222, Ap
c	11	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	12	32.8	7.2	24081	14	US-10-132-134-13	Sequence 13, Appl
	13	32.8	7.2	52101	14	US-10-132-134-1	Sequence 1, Appli
c	14	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
	15	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
	16	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
	17	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c	18	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
c	19	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
c	20	31.4	6.9	460	15	US-10-260-238-787	Sequence 787, App
	21	31.4	6.9	7374	10	US-09-764-891-10207	Sequence 10207, A
	22	31.4	6.9	7374	14	US-10-205-428-1006	Sequence 1006, Ap
c	23	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
	24	31.2	6.8	2266	15	US-10-108-260A-375	Sequence 375, App
	25	31	6.8	1049	12	US-10-142-426-358	Sequence 358, App
	26	31	6.8	1049	14	US-10-123-155-358	Sequence 358, App
	27	31	6.8	1049	14	US-10-146-731-358	Sequence 358, App
	28	31	6.8	1049	14	US-10-140-472-358	Sequence 358, App
	29	31	6.8	1049	14	US-10-141-761-358	Sequence 358, App
	30	31	6.8	1049	14	US-10-142-885-358	Sequence 358, App
	31	31	6.8	1049	14	US-10-158-790-358	Sequence 358, App
	32	31	6.8	1049	15	US-10-137-871-358	Sequence 358, App
	33	31	6.8	1049	15	US-10-140-923-358	Sequence 358, App
	34	31	6.8	1049	15	US-10-141-756-358	Sequence 358, App
	35	31	6.8	1049	15	US-10-141-759-358	Sequence 358, App
	36	31	6.8	1049	15	US-10-140-805-358	Sequence 358, App
	37	31	6.8	1049	15	US-10-140-864-358	Sequence 358, App
c	38	31	6.8	1841	15	US-10-369-493-45479	Sequence 45479, A
c	39	30.8	6.8	524	14	US-10-029-386-4400	Sequence 4400, Ap
	40	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
c	41	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
	42	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
	43	30.6	6.7	1421	12	US-10-425-114-12527	Sequence 12527, A
c	44	30.6	6.7	1573	9	US-09-771-161A-1	Sequence 1, Appli
	45	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1662)
US-10-156-761-3483

Query Match 7.6%; Score 34.6; DB 14; Length 1662;
Best Local Similarity 49.7%; Pred. No. 0.79;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
Db 590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531

```
Qy        110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
Db        530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCC 226
Db        470 TTGAACTTCTTGGCGGCGACGGCGGCGAACGTCTCGTACGTGCCCTGGAGGATGCCC 414
```

RESULT 2
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 7.6%; Score 34.6; DB 14; Length 9025608;
Best Local Similarity 49.7%; Pred. No. 77;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy         50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
Db    4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

Qy        110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
Db    4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCC 226
Db    4331901 TTGAACTTCTTGGCGGCGACGGCGGCGAACGTCTCGTACGTGCCCTGGAGGATGCCC 4331957
```

RESULT 3
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1.1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
Db         89 GC 88
```

RESULT 4
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1.1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
Db         89 GC 88
```

RESULT 5
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF

```
;  FILE REFERENCE: 21829/81
;  CURRENT APPLICATION NUMBER: US/09/879,248
;  CURRENT FILING DATE:  2001-06-12
;  PRIOR APPLICATION NUMBER: 60/212,211
;  PRIOR FILING DATE: 2000-06-16
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-879-248-5

  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 1.1;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |     |    ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 6
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Leon, Ernesto
;  APPLICANT:  Oviedo, Agustin
;  TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
;  TITLE OF INVENTION:  FROM ORNAMENTAL PLANTS
;  FILE REFERENCE: 21829/111
;  CURRENT APPLICATION NUMBER: US/10/010,390
;  CURRENT FILING DATE:  2001-11-05
;  PRIOR APPLICATION NUMBER: 60/248,169
;  PRIOR FILING DATE: 2000-11-13
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-010-390-6

  Query Match             7.5%;  Score 34;  DB 14;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 1.1;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |     |    ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 7
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION:  RESISTANCE
;  FILE REFERENCE: 21829/203 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/10/441,736
;  CURRENT FILING DATE:  2003-05-20
;  PRIOR APPLICATION NUMBER: 60/107,243
;  PRIOR FILING DATE: 1998-11-05
;  PRIOR APPLICATION NUMBER: 09/431,614
;  PRIOR FILING DATE: 1999-11-02
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-441-736-5

  Query Match             7.5%;  Score 34;  DB 15;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 1.1;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     |||||  | |  ||   |     |    ||| | | |||||  |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 8
US-10-184-644-418/c
; Sequence 418, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C227
;  CURRENT APPLICATION NUMBER: US/10/184,644
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 418
;   LENGTH: 655
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-644-418

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 655;
  Best Local Similarity   14.9%;  Pred. No. 1.2;
  Matches   39;  Conservative   76;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         18 GGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATG 77
              :::  |||  :   :  ::::    : ::      : ::  |: ::  : : |      :
```

```
Db        325 SSM.CCAT.KS...M..DYDD.C..DNBB..N..AMNBH.MCYTSSANH.BAC...SS.B 266

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137

Db        265 M.YBSTSSNS.ASNKCM.BRASS.B.DA.KD..Y...TAC.S.SASSS.S..ACGBNHA. 206

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATA 197

Db        205 MAC.M.BB.N.S.GHARTMGMKBSS.BHS.ACYTG.MGYBH.A.ACMMYBCWCB.GBAD. 146

Qy        198 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGA 257

Db        145 TGATNS..KC..GAG.YHA.TTG..M..K.W.G..SGHDGM..CN.DYA.ACB.GSSGBY 86

Qy        258 TAATAGTACTATTGAACTGACT 279

Db         85 .SANAGD.SBRACT.GMHGA.B 64


RESULT 9
US-10-184-634-418/c
; Sequence 418, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C217
;  CURRENT APPLICATION NUMBER: US/10/184,634
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 418
;   LENGTH: 655
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-634-418

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 655;
  Best Local Similarity   14.9%;  Pred. No. 1.2;
  Matches   39;  Conservative   76;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         18 GGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATG 77

Db        325 SSM.CCAT.KS...M..DYDD.C..DNBB..N..AMNBH.MCYTSSANH.BAC...SS.B 266

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137

Db        265 M.YBSTSSNS.ASNKCM.BRASS.B.DA.KD..Y...TAC.S.SASSS.S..ACGBNHA. 206

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATA 197

Db        205 MAC.M.BB.N.S.GHARTMGMKBSS.BHS.ACYTG.MGYBH.A.ACMMYBCWCB.GBAD. 146

Qy        198 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGA 257

Db        145 TGATNS..KC..GAG.YHA.TTG..M..K.W.G..SGHDGM..CN.DYA.ACB.GSSGBY 86

Qy        258 TAATAGTACTATTGAACTGACT 279

Db         85 .SANAGD.SBRACT.GMHGA.B 64


RESULT 10
US-10-156-761-7222/c
; Sequence 7222, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7222
;   LENGTH: 1848
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1848)
US-10-156-761-7222

  Query Match             7.3%;  Score 33.2;  DB 14;  Length 1848;
  Best Local Similarity   47.6%;  Pred. No. 2.5;
  Matches   98;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105

Db        834 GGCGGCGTACACCGCGGCAGCGAGGCCGCCGGGTCCGGAGCCGACGATGAGGACGTCGCC 775

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165

Db        774 GTGACGGTTTCCGGCGGGGAGTGACGGCAGGCCGATGAGCCGGGCCAGTTCCGCGTTGCT 715

Qy        166 GCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACC 225

Db        714 GGGGTTGCGCAACAGGGTGGTGTCCCGCCAGATGACGAGCGGTGTCTCCTCCGGGCCGAC 655

Qy        226 CATGAAATGGCACATGCAGGCCAGGG 251

Db        654 GCCGAAGCGGCGCAGCAGGTCCTCGG 629


RESULT 11
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
```

```
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             7.3%;  Score 33.2;  DB 14;  Length 9025608;
  Best Local Similarity   47.6%;  Pred. No. 1.9e+02;
  Matches   98;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105
              ||| | |  |     |||  ||   |     ||  ||| | ||   || |  ||| || |
Db    8638225 GGCGGCGTACACCGCGGCAGCGAGGCCGCCGGGTCCGGAGCCGACGATGAGGACGTCGCC 8638166

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165
              |  |    |||||||  ||| | |  |  |  |  | | | |  |   | |      |||
Db    8638165 GTGACGGTTTCCGGCGGGGAGTGACGGCAGGCCGATGAGCCGGGCCAGTTCCGCGTTGCT 8638106

Qy        166 GCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACC 225
              | | |||| |  ||  | | || ||    | ||||  ||| |  |  ||      |    |
Db    8638105 GGGGTTGCGCAACAGGGTGGTGTCCCGCCAGATGACGAGCGGTGTCTCCTCCGGGCCGAC 8638046

Qy        226 CATGAAATGGCACATGCAGGCCAGGG 251
                |||  ||| ||    | ||  ||
Db    8638045 GCCGAAGCGGCGCAGCAGGTCCTCGG 8638020


RESULT 12
US-10-132-134-13
; Sequence 13, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Yang, Xianshu
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
;  FILE REFERENCE: 3012-2US
;  CURRENT APPLICATION NUMBER: US/10/132,134
;  CURRENT FILING DATE:  2002-04-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
;   LENGTH: 24081
;   TYPE: DNA
;   ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-13

  Query Match             7.2%;  Score 32.8;  DB 14;  Length 24081;
  Best Local Similarity   50.0%;  Pred. No. 14;
  Matches   82;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105
              ||| |  | ||||  | || |   || |  |||| || |||||| |        | || |
Db      20227 GGCCGCACACTGGTGGTCGGCTGGCCCCTGTGGGCCGACGGCGGCATGACGGTGGACGCC 20286

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165
              | ||       | ||| | ||   |   |  ||||    | |  | |  || |  ||| |
Db      20287 GCCACCGAACGCCGCCTGCACGACAGCGTCGGCATGGTGCCGATCCGCGCCCCGCACGGT 20346

Qy        166 GCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGT 209
              | |   || |||| | |||     | | |      || | ||  ||
Db      20347 GTGGAGGCGCTGCTACGCGCCTACGGCACCGGCGACCCGCACGT 20390


RESULT 13
US-10-132-134-1
; Sequence 1, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Yang, Xianshu
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
;  FILE REFERENCE: 3012-2US
;  CURRENT APPLICATION NUMBER: US/10/132,134
;  CURRENT FILING DATE:  2002-04-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 52101
;   TYPE: DNA
;   ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-1

  Query Match             7.2%;  Score 32.8;  DB 14;  Length 52101;
  Best Local Similarity   50.0%;  Pred. No. 21;
  Matches   82;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105
              ||| |  | ||||  | || |   || |  |||| || |||||| |        | || |
Db      37860 GGCCGCACACTGGTGGTCGGCTGGCCCCTGTGGGCCGACGGCGGCATGACGGTGGACGCC 37919

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165
              | ||         | ||| | ||    |   |  ||||    | |  | |  || |  ||| |
Db      37920 GCCACCGAACGCCGCCTGCACGACAGCGTCGGCATGGTGCCGATCCGCGCCCCGCACGGT 37979

Qy        166 GCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGT 209
              | |   || |||| | |||     | | |      || | ||  ||
Db      37980 GTGGAGGCGCTGCTACGCGCCTACGGCACCGGCGACCCGCACGT 38023


RESULT 14
US-10-369-493-33133/c
; Sequence 33133, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33133
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Xylella fastidiosa
US-10-369-493-33133

  Query Match             7.1%;  Score 32.2;  DB 15;  Length 1083;
  Best Local Similarity   59.1%;  Pred. No. 4.1;
  Matches   55;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTG 421
              ||  | ||| |||||    |  | |||||     ||  || | ||| | ||| ||||
Db        488 ATCTCACCGAGCTGGGATAAAAAACCGCGCACAATCACACCATAACGCTGCTTCAGCCAC 429

Qy        422 GTTTTGGCAACAACGCCACGGCTAACCAGTATT 454
                 |||| ||||| || |||     | ||  | |||
Db        428 TTTTTAGCAACCACCCCAGCAGTCACACGCATT 396


RESULT 15
US-09-815-242-6028
; Sequence 6028, Application US/09815242
```

```
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6028
;   LENGTH: 930
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(930)
US-09-815-242-6028

  Query Match             7.0%;  Score 31.8;  DB 9;  Length 930;
  Best Local Similarity   52.7%;  Pred. No. 5.2;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              | |||  |  || ||||    | |  || |  | ||| ||   |   ||  |  || |||
Db        233 AGCAGCAGGGCGGTAAAGCACCGGCAATGATGGCCGGTCACAGCCTGGGGGAATACTCCG 292

Qy        371 CGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCA 430
              ||||||||          |    ||||| |   |    |||||||| |||||   |  ||
Db        293 CGCTGGTTTGCGCTGGTGTGATTGATTTCGCTGATGCGGTGCGTCTGGTTGAGATGCGCG 352

Qy        431 ACAACGCCACG 441
              |||   || |
Db        353 GCAAGTTCATG 363


Search completed: March 19, 2004, 03:50:20
Job time : 276.734 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-19	Sequence 19, Appl
2	412.8	90.5	456	6	US-08-233-642A-56	Sequence 56, Appl
3	412.8	90.5	456	23	US-09-543-407-1	Sequence 1, Appli
4	385.6	84.6	456	23	US-09-543-407-25	Sequence 25, Appl
5	363.2	79.6	456	23	US-09-543-407-17	Sequence 17, Appl
6	361.6	79.3	456	23	US-09-543-407-11	Sequence 11, Appl
7	361.6	79.3	456	23	US-09-543-407-13	Sequence 13, Appl
8	361.6	79.3	456	23	US-09-543-407-23	Sequence 23, Appl
9	360	78.9	456	23	US-09-543-407-15	Sequence 15, Appl
10	360	78.9	456	23	US-09-543-407-27	Sequence 27, Appl
11	358.4	78.6	456	23	US-09-543-407-21	Sequence 21, Appl
12	355.2	77.9	456	23	US-09-543-407-29	Sequence 29, Appl
13	288.6	63.3	361	6	US-08-233-642A-54	Sequence 54, Appl
14	236.8	51.9	456	23	US-09-543-407-3	Sequence 3, Appli
15	236.8	51.9	675	13	US-08-978-878-3	Sequence 3, Appli
16	236.8	51.9	675	31	US-09-741-873B-3	Sequence 3, Appli
17	234.4	51.4	453	92	US-60-352-946-1	Sequence 1, Appli
18	234.4	51.4	453	101	US-60-444-371-1	Sequence 1, Appli
19	227.2	49.8	477	17	US-09-252-691-172	Sequence 172, App
20	227.2	49.8	477	17	US-09-252-691C-172	Sequence 172, App
21	227.2	49.8	477	51	US-10-417-886-172	Sequence 172, App
22	192.8	42.3	396	13	US-08-978-878-1	Sequence 1, Appli
23	168	36.8	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
C 28	50	11.0	78	23	US-09-543-407-52	Sequence 52, Appl
29	50	11.0	78	23	US-09-543-407-53	Sequence 53, Appl
30	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
31	45	9.9	78	23	US-09-543-407-49	Sequence 49, Appl
C 32	43.4	9.5	78	23	US-09-543-407-48	Sequence 48, Appl
33	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
C 34	42	9.2	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
C 35	36.8	8.1	1290	37	US-09-902-540-5006	Sequence 5006, Ap
C 36	36.8	8.1	12950	37	US-09-902-540-1036	Sequence 1036, Ap
C 37	36.2	7.9	632	33	US-09-815-264-23539	Sequence 23539, A
C 38	36.2	7.9	805	26	US-09-620-392-61621	Sequence 61621, A
C 39	36.2	7.9	805	30	US-09-702-134-34173	Sequence 34173, A
C 40	36.2	7.9	805	33	US-09-815-264-61323	Sequence 61323, A
41	35.8	7.9	387	11	US-08-790-774-9199	Sequence 9199, Ap
42	35.8	7.9	387	37	US-09-912-293-217377	Sequence 217377,
C 43	35.8	7.9	1281	20	US-09-417-507-7264	Sequence 7264, Ap
C 44	35	7.7	2256646	52	US-10-470-565-1	Sequence 1, Appli
C 45	34.8	7.6	542	24	US-09-565-306-14113	Sequence 14113, A

ALIGNMENTS

```
RESULT 1
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-136;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-642A-56

  Query Match             90.5%;  Score 412.8;  DB 6;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 2.1e-122;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             90.5%;  Score 412.8;  DB 23;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 2.1e-122;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             84.6%;  Score 385.6;  DB 23;  Length 456;
  Best Local Similarity   90.4%;  Pred. No. 1.4e-113;
  Matches  412;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 79.6%; Score 363.2; DB 23; Length 456;
Best Local Similarity 87.3%; Pred. No. 2.5e-106;
Matches 398; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 8.2e-106;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             79.3%;  Score 361.6;  DB 23;  Length 456;
  Best Local Similarity   87.1%;  Pred. No. 8.2e-106;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  ||||||  | |  | |  ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||         |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             79.3%;  Score 361.6;  DB 23;  Length 456;
  Best Local Similarity   87.1%;  Pred. No. 8.2e-106;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  ||||||  | |  | |  ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||  |||     ||  ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             78.9%;  Score 360;  DB 23;  Length 456;
  Best Local Similarity   86.8%;  Pred. No. 2.7e-105;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||  ||    | | | |   | |     |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             78.9%;  Score 360;  DB 23;  Length 456;
  Best Local Similarity   86.8%;  Pred. No. 2.7e-105;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||  |         | | ||  |     |       ||| ||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             78.6%;  Score 358.4;  DB 23;  Length 456;
  Best Local Similarity   86.6%;  Pred. No. 8.9e-105;
  Matches  395;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |     | |   ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||   |  |      | ||  |   ||  |   |   | ||
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 9.7e-104;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |     | |   ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||   || |||  |    |                    |  | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 63.3%; Score 288.6; DB 6; Length 361;
Best Local Similarity 91.3%; Pred. No. 3.4e-82;
Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              ||||||||||||| |||   |  ||||||    | |     | |     ||  ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. Coli
US-09-543-407-3

Query Match 51.9%; Score 236.8; DB 23; Length 456;
Best Local Similarity 70.0%; Pred. No. 2.3e-65;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||| ||    || |||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  ||||||||||| | |||   |    ||||||    ||    | | |    ||| ||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| ||||||  ||| |  |||  ||||||| |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

RESULT 15
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-08-978-878-3

Query Match 51.9%; Score 236.8; DB 13; Length 675;
Best Local Similarity 70.0%; Pred. No. 2.6e-65;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||| ||    || |||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  ||||||||||| | |||   |    ||||||    ||    | | |    ||| ||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| ||||||  ||| |  |||  ||||||| |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

Search completed: March 19, 2004, 03:01:58
Job time : 2652.77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	236.8	51.9	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	192.8	42.3	396	5	US-09-741-873C-1	Sequence 1, Appli
c	3	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
	4	30.4	6.7	848	6	US-10-100-683-643	Sequence 643, App
	5	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
c	6	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	7	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
c	8	29	6.4	1046	6	US-10-048-770C-3	Sequence 3, Appli
c	9	28.8	6.3	3816	1	PCT-US04-04280-78	Sequence 78, Appl
c	10	28.8	6.3	4333	6	US-10-152-372-233	Sequence 233, App
c	11	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	12	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
c	13	28.6	6.3	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	14	28.6	6.3	3416	6	US-10-100-683-3397	Sequence 3397, Ap
c	15	28.6	6.3	8206	6	US-10-021-698A-3529	Sequence 3529, Ap
	16	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	17	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
	18	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	19	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	20	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
c	21	28.2	6.2	14449	6	US-10-021-698A-924	Sequence 924, App
	22	28.2	6.2	225724	6	US-10-021-698A-711	Sequence 711, App
	23	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
c	24	28	6.1	554	6	US-10-767-701-27293	Sequence 27293, A
	25	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
c	26	27.8	6.1	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	27	27.8	6.1	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	28	27.8	6.1	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	29	27.8	6.1	1400	7	US-60-545-213-6407	Sequence 6407, Ap
	30	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
c	31	27.6	6.1	363	6	US-10-767-701-29572	Sequence 29572, A
	32	27.6	6.1	463	6	US-10-767-701-17923	Sequence 17923, A
	33	27.6	6.1	529	6	US-10-767-701-19620	Sequence 19620, A
	34	27.6	6.1	551	6	US-10-767-701-3483	Sequence 3483, Ap
	35	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
	36	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
	37	27.6	6.1	617	6	US-10-781-469-57	Sequence 57, Appl
	38	27.6	6.1	632	6	US-10-100-683-3854	Sequence 3854, Ap
c	39	27.6	6.1	946	6	US-10-784-004-232	Sequence 232, App
	40	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
c	41	27.6	6.1	1818	6	US-10-773-236-118	Sequence 118, App
	42	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	43	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	44	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	45	27.6	6.1	8654	6	US-10-021-698A-3499	Sequence 3499, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 51.9%; Score 236.8; DB 5; Length 675;
Best Local Similarity 70.0%; Pred. No. 3.4e-72;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 42.3%; Score 192.8; DB 5; Length 396;
Best Local Similarity 67.9%; Pred. No. 4.5e-57;
Matches 269; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.2;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376

Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422

Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583
```

RESULT 4
US-10-100-683-643
; Sequence 643, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314

```
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 643
;   LENGTH: 848
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-643

  Query Match           6.7%;  Score 30.4;  DB 6;  Length 848;
  Best Local Similarity   49.4%;  Pred. No. 1.2;
  Matches   79;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              ||| ||  | | ||  | ||   |  |||||     |      |   ||  |||
Db          1 GGCACGAGCACTACTGTAAGAGCTGGTCAGTGAATGTGGTTGCAGCATGGCCTTTGGGCA 60

Qy        178 CAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 237
              |||  |  |   ||| ||  | |||||||||  ||    |   |   || ||| ||
Db         61 AGAAGTAACCCATTTAACTAAAACCAGCTGGTTGGCCCCACTCAGATTTATCAAAGGGTT 120

Qy        238 CATGCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGA 277
              ||    || ||| | ||||| |  |  ||||  ||| |
Db        121 ACTGGGTCCCTGGGGGTGGATATTGCTTATATTAGACTTA 160


RESULT 5
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match           6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 1.6;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||    |       ||||| |  |     | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
              |||||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 6
US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-117

  Query Match           6.4%;  Score 29.2;  DB 6;  Length 249;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||       || || ||| |   | | | |   || | | |||
Db        180 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 121

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    | | ||  || || || ||| |  | |||    |||| | | |    || |
Db        120 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 61

Qy        415 CA 416
              |
Db         60 GA 59


RESULT 7
US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 757
;   TYPE: DNA
;   ORGANISM: E. Coli
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(757)
;   OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

  Query Match           6.4%;  Score 29.2;  DB 6;  Length 757;
  Best Local Similarity   52.5%;  Pred. No. 2.9;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||       || || ||| |   | | | |   || | | |||
Db        264 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 323

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    | | ||  || || || ||| |  | |||    |||| | | |    || |
Db        324 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 383

Qy        415 CA 416
              |
```

```
Db        384 GA 385

RESULT 8
US-10-048-770C-3/c
; Sequence 3, Application US/10048770C
; GENERAL INFORMATION:
;  APPLICANT: NEW BIOTECHNIC, S.A.
;  APPLICANT:  UNIVERSIDAD DE SEVILLA
;  TITLE OF INVENTION: ANTI-FUNGAL COMPOSITIONS AND METHOD FOR CONTROLLING FUNGI
;  FILE REFERENCE: 1837-2
;  CURRENT APPLICATION NUMBER: US/10/048,770C
;  CURRENT FILING DATE:  2002-01-31
;  PRIOR APPLICATION NUMBER: PCT/ES00/00292
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 1046
;   TYPE: DNA
;   ORGANISM: Trichoderma harzianum
US-10-048-770C-3

  Query Match             6.4%;  Score 29;  DB 6;  Length 1046;
  Best Local Similarity   49.0%;  Pred. No. 4;
  Matches   77;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGT 216
              || |||| ||     |  |||  | | ||  |||| | | | |  |||||  ||
Db        760 GCAAACGTTGGTGATATCCATCAATGAGGTTCCGCTAAAGACGGTCTGACCGTTACGAGT 701

Qy        217 GTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
                  | | |    |||| |||| |||| || |  |||  | | |     |  |
Db        700 CAGAGCAAACTTTGCTGGCCGATGCTGGCCGGGATTTGGAAGAGATTGGCTCCGGCGTT 641

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACC 313
              | | |  |||| ||   |  |||   ||||
Db        640 GGCCTGGAACGTTTGAGTGGTGCCACCAGACGTGACC 604

RESULT 9
PCT-US04-04280-78/c
; Sequence 78, Application PC/TUS0404280
; GENERAL INFORMATION:
;  APPLICANT: INCYTE CORPORATION; KABLE, Amy E.;
;  APPLICANT:  YUE, Henry; BAUGHN, Mariah R.;
;  APPLICANT:  TRIBOULEY, Catherine M.; RING, Huijun Z.;
;  APPLICANT:  TRAN, Uyen K.; EMERLING, Brooke E.;
;  APPLICANT:  RAMKUMAR, Jayalaxmi; HAFALIA, April J.A.;
;  APPLICANT:  SWARNAKAR, Anita; LEE, Soo Yeun;
;  APPLICANT:  CHAWLA, Narinder K.; GIETZEN, Kimberly J.;
;  APPLICANT:  MARQUIS, Joseph P.; ELLIOTT, Vicki S.;
;  APPLICANT:  BECHA, Shanya D.; FAVERO, Kristin D.;
;  APPLICANT:  WANG, Jonathan T.; NAIDU, Sangeeta;
;  APPLICANT:  HAWKINS, Phillip R.; JIN, Pei;
;  APPLICANT:  CHIEN, David
;  TITLE OF INVENTION: ENZYMES
;  FILE REFERENCE: PF-1507 PCT
;  CURRENT APPLICATION NUMBER: PCT/US04/04280
;  CURRENT FILING DATE:  2004-02-12
;  PRIOR APPLICATION NUMBER: US 60/447,246
;  PRIOR FILING DATE: 2003-02-12
;  PRIOR APPLICATION NUMBER: US 60/449,087
;  PRIOR FILING DATE: 2003-02-21
;  PRIOR APPLICATION NUMBER: US 60/450,622
;  PRIOR FILING DATE: 2003-02-26
;  PRIOR APPLICATION NUMBER: US 60/456,704
;  PRIOR FILING DATE: 2003-03-21
;  PRIOR APPLICATION NUMBER: US 60/463,194
;  PRIOR FILING DATE: 2003-04-15
;  PRIOR APPLICATION NUMBER: US 60/469,358
;  PRIOR FILING DATE: 2003-05-09
;  PRIOR APPLICATION NUMBER: US 60/475,532
;  PRIOR FILING DATE: 2003-06-02
;  PRIOR APPLICATION NUMBER: US 60/476,278
;  PRIOR FILING DATE: 2003-06-04
;  PRIOR APPLICATION NUMBER: US 60/483,395
;  PRIOR FILING DATE: 2003-06-27
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: PERL Program
; SEQ ID NO 78
;   LENGTH: 3816
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 7517280CB1
PCT-US04-04280-78

  Query Match             6.3%;  Score 28.8;  DB 1;  Length 3816;
  Best Local Similarity   50.7%;  Pred. No. 8.8;
  Matches   69;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              ||||  |||  |   ||||  |  ||  | | ||  |  |||  ||     |||
Db       2245 GGCCATGACAAAGGTATGAGGAAAGAGCATAGCTGCACCCTTCCTCATGCCCTGAAGCTG 2186

Qy        178 CAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 237
              || || | |||||   || ||||     ||  | | |||  |  ||||  |     |
Db       2185 CACAGTGTTGCCCTGGAACATGACTCCATGCACATCAGTCTCTGTGCCCAGGCCGAGCAG 2126

Qy        238 CATGCAGGCCAGGGTG 253
              |||||||  |||
Db       2125 GTGCCAGGCCACTGTG 2110

RESULT 10
US-10-152-372-233/c
; Sequence 233, Application US/10152372
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C394
;  CURRENT APPLICATION NUMBER: US/10/152,372
;  CURRENT FILING DATE:  2002-05-21
;  PRIOR APPLICATION NUMBER: 60/049911
;  PRIOR FILING DATE: 1997-06-18
;  PRIOR APPLICATION NUMBER: 60/056974
;  PRIOR FILING DATE: 1997-08-26
;  PRIOR APPLICATION NUMBER: 60/059113
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059115
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059117
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059122
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059184
;  PRIOR FILING DATE: 1997-09-17
```

```
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059352
;  PRIOR FILING DATE: 1997-09-19
;  PRIOR APPLICATION NUMBER: 60/059588
;  PRIOR FILING DATE: 1997-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 233
;   LENGTH: 4333
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-152-372-233

  Query Match             6.3%;  Score 28.8;  DB 6;  Length 4333;
  Best Local Similarity   50.7%;  Pred. No. 9.4;
  Matches   69;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              |||| ||| |   |||| |  | ||  | |  ||  |     ||| ||    |||
Db       2191 GGCCATGACAAAGGTATGAGGAAAGAGCATAGCTGCACCCTTCCTCATGCCCTGAAGCTG 2132

Qy        178 CAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 237
              || || | |||||   || ||||     ||   | | || |   | |||| |    |
Db       2131 CACAGTGTTGCCCTGGAACATGACTCCATGCACATCAGTCTCTGTGCCCAGGCCGAGCAG 2072

Qy        238 CATGCAGGCCAGGGTG 253
                |||||||  |||
Db       2071 GTGCCAGGCCACTGTG 2056


RESULT 11
US-10-417-375A-143/c
; Sequence 143, Application US/10417375A
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375A
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143
;   LENGTH: 83009
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(83009)
;   OTHER INFORMATION: n = A,T,C or G
US-10-417-375A-143

  Query Match             6.3%;  Score 28.8;  DB 6;  Length 83009;
  Best Local Similarity   48.8%;  Pred. No. 40;
  Matches   78;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        281 AGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTA 340
              |||||  |   ||||   ||   | | || |  ||   |  |   | | |||  ||||
Db      14056 AGAATATTAAAAGAATGCATAATATCCTCAAAAAGAACATTGACTAGTATTCCTTTATTC 13997

Qy        341 CTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCA 400
              || ||    |    |  |||||    |    |       || ||| |  |   |  ||| |
Db      13996 CTTTCAAGTACACTTAAGTTAATGCTTCAATGAAATACAAGCAGGCACCTATTACTTCTA 13937

Qy        401 GCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCAC 440
              |  |||||   | ||||  | | ||   |  || | |||||
Db      13936 GAATAATGAGACCTCAGATGGATTACAGGCACGATGCCAC 13897


RESULT 12
PCT-US04-05654-1030
; Sequence 1030, Application PC/TUS0405654
; GENERAL INFORMATION:
;  APPLICANT: Sherman, Bradley K
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline  E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Dubell III, Arnold N
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Repetti, Peter
;  APPLICANT:  Century, Karen
;  APPLICANT:  Gutterson, Neal
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Kumimoto, Roderick W
;  APPLICANT:  Pilgrim, Marsha L
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 PCT
;  CURRENT APPLICATION NUMBER: PCT/US04/05654
;  CURRENT FILING DATE:  2004-03-03
;  PRIOR APPLICATION NUMBER: 10/374,780
;  PRIOR FILING DATE: 2003-02-25
;  PRIOR APPLICATION NUMBER: 10/675,852
;  PRIOR FILING DATE: 2003-09-30
;  NUMBER OF SEQ ID NOS: 2950
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1030
;   LENGTH: 1186
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
PCT-US04-05654-1030

  Query Match             6.3%;  Score 28.6;  DB 1;  Length 1186;
  Best Local Similarity   59.0%;  Pred. No. 5.9;
  Matches   49;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         44 CTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCG 103
              | || | |||||  || | || ||       |  | || ||||||||   |  |   ||||
Db        606 CAGGGACTGCTCCAGCGGCCGGCGGCGGCTACGGCGGTGGCGGCGGCGGTGGTGGTGGCG 665

Qy        104 GCGGCAATAGTTCCGGCCCGGAC 126
              || || | |  | |||||  || |
Db        666 GCTGCTACAACTGCGGCCAGGCC 688


RESULT 13
US-10-100-683-5480/c
; Sequence 5480, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
```

```
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5480
;   LENGTH: 1295
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-5480

  Query Match             6.3%;  Score 28.6;  DB 6;  Length 1295;
  Best Local Similarity   49.7%;  Pred. No. 6.1;
  Matches   73;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        242 CAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
              | | ||  | |||   |||    |  | | | ||| ||     ||       |   | ||
Db        767 CGGACCCTGCTGCCCCTAACCTCAAGAGTCAGCTGGCTGCAGCTGCCAGGGCAGTAACTG 708

Qy        302 CCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
              || |||| | |||   | | |||    | || |      |   ||||||  | |   ||
Db        707 ACAGCATCAATCAGCTCATCACTATGTGCACCCAGCAGGCACCCGGCCAGAAGGAGTGTG 648

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCG 388
              ||||||||  || ||      |||| |
Db        647 ATAACGCCCTGCGGGAATTGGAGACGG 621


RESULT 14
US-10-100-683-3397/c
; Sequence 3397, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3397
;   LENGTH: 3416
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-3397

  Query Match             6.3%;  Score 28.6;  DB 6;  Length 3416;
  Best Local Similarity   49.7%;  Pred. No. 9.8;
  Matches   73;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        242 CAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
              | | ||  | |||   |||    |  | | | ||| ||     ||       |   | ||
Db        761 CGGACCCTGCTGCCCCTAACCTCAAGAGTCAGCTGGCTGCAGCTGCCAGGGCAGTAACTG 702

Qy        302 CCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
              || |||| | |||   | | |||    | || |      |   ||||||  | |   ||
Db        701 ACAGCATCAATCAGCTCATCACTATGTGCACCCAGCAGGCACCCGGCCAGAAGGAGTGTG 642

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCG 388
              ||||||||  || ||      |||| |
Db        641 ATAACGCCCTGCGGGAATTGGAGACGG 615


RESULT 15
US-10-021-698A-3529/c
; Sequence 3529, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 3529
;   LENGTH: 8206
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (401)
;   OTHER INFORMATION: a, t, c or g
US-10-021-698A-3529

  Query Match             6.3%;  Score 28.6;  DB 6;  Length 8206;
  Best Local Similarity   49.7%;  Pred. No. 15;
  Matches   73;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         242 CAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
               | | ||  | |||   |||    |  | | | ||| ||     ||       |   | ||
Db        4174 CGGACCCTGCTGCCCCTAACCTCAAGAGTCAGCTGGCTGCAGCTGCCAGGGCAGTAACTG 4115

Qy         302 CCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
               || |||| | |||   | | |||    | || |      |   ||||||  | |   ||
Db        4114 ACAGCATCAATCAGCTCATCACTATGTGCACCCAGCAGGCACCCGGCCAGAAGGAGTGTG 4055

Qy         362 ATAACGCCGCGCTGGTTAATCAGACCG 388
               ||||||||  || ||      |||| |
Db        4054 ATAACGCCCTGCGGGAATTGGAGACGG 4028


Search completed: March 19, 2004, 03:10:16
Job time : 55.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	237.4	52.1	680	12	BJ618688	BJ618688 BJ618688
c	2	37.4	8.2	484	9	AV430994	AV430994 AV430994
c	3	35.4	7.8	664	13	CA141136	CA141136 SCJFRT205
	4	34.8	7.6	675	13	BX251172	BX251172 BX251172
c	5	34.6	7.6	699	12	BI959328	BI959328 HVSMEn001
	6	34.2	7.5	296	9	AV058630	AV058630 AV058630
c	7	34.2	7.5	669	28	AZ959284	AZ959284 2M0226G20
	8	34.2	7.5	978	28	BZ569281	BZ569281 pacs2-164
	9	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	10	34.2	7.5	1137	28	BZ565783	BZ565783 pacs2-164
c	11	34	7.5	234	10	BB581987	BB581987 BB581987
c	12	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
	13	34	7.5	500	28	AZ180696	AZ180696 SP_0181_A
c	14	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	15	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	16	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
c	17	33.8	7.4	902	28	BZ569250	BZ569250 pacs2-164
	18	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	19	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	20	33.6	7.4	680	10	BF668888	BF668888 602123291
	21	33.6	7.4	966	13	BQ718166	BQ718166 AGENCOURT
c	22	33.4	7.3	505	6	AL809250	Al809250 Triticum
	23	33.4	7.3	523	12	BJ334624	BJ334624 BJ334624
	24	33.4	7.3	715	13	BU444683	BU444683 603765715
	25	33.4	7.3	1064	13	BQ939900	BQ939900 AGENCOURT
c	26	33.2	7.3	317	12	BM333324	BM333324 MEST189-B
	27	33.2	7.3	625	10	BF501499	BF501499 AT16857.5
	28	33.2	7.3	629	9	AU295325	AU295325 AU295325
c	29	33.2	7.3	806	13	BX888120	BX888120 BX888120
	30	33	7.2	509	12	BM896625	BM896625 ph53g06.y
	31	33	7.2	649	12	BI261742	BI261742 602954071
c	32	33	7.2	826	28	BZ462870	BZ462870 BONBW34TR
c	33	32.8	7.2	521	14	CA703231	CA703231 wdk1c.pk0
	34	32.8	7.2	572	12	BI626794	BI626794 RH67848.5
	35	32.8	7.2	594	12	BJ335653	BJ335653 BJ335653
	36	32.8	7.2	613	14	CB935126	CB935126 ru03d09.y
	37	32.8	7.2	653	12	BJ336964	BJ336964 BJ336964
	38	32.8	7.2	1893	29	AY414592	AY414592 Pan trogl
	39	32.6	7.1	204	28	AZ577991	AZ577991 17d03 Sho
c	40	32.6	7.1	352	9	AA784257	AA784257 d3g05a1.f
c	41	32.6	7.1	356	9	AA785291	AA785291 g6c08a1.f
	42	32.6	7.1	778	29	CNS037J5	AL231386 Tetraodon
	43	32.6	7.1	813	28	BZ204853	BZ204853 CH230-390
	44	32.6	7.1	919	14	CK246169	CK246169 EST729806
	45	32.6	7.1	942	14	CK247920	CK247920 EST731557

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

```
          http://xenopus.nibb.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match             52.1%;  Score 237.4;  DB 12;  Length 680;
  Best Local Similarity   70.0%;  Pred. No. 6.4e-57;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||   | ||    | ||| |||||| |||||||||  || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  |||||||||| | |||    |  ||||||    ||    | | |   ||| ||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138


RESULT 2
AV430994/c
LOCUS       AV430994                484 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PL027e10_r 5', mRNA sequence.
ACCESSION   AV430994
VERSION     AV430994.1  GI:8586219
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
   PUBMED   10907854
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Porphyra yezoensis"
                     /mol_type="mRNA"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PL027e10_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             8.2%;  Score 37.4;  DB 9;  Length 484;
  Best Local Similarity   57.1%;  Pred. No. 10;
  Matches   68;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |   | |  | | ||  ||| |||||   ||| ||||| ||||||  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
                | ||  ||   ||  ||||  ||| ||||||     |  ||     |||  |  || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323


RESULT 3
CA141136/c
LOCUS       CA141136                664 bp    mRNA    linear   EST 24-SEP-2003
DEFINITION  SCJFRT2058D10.g RT2 Saccharum officinarum cDNA clone SCJFRT2058D10
            5', mRNA sequence.
ACCESSION   CA141136
VERSION     CA141136.1  GI:35034821
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 058  row: D  column: 10
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCJFRT2058D10"
                     /lab_host="DH10B"
                     /clone_lib="RT2"
                     /note="Organ: Root tips(0.3cm-long) from adult plants;
                     Vector: pSport1; Site_1: SalI; Site_2: NotI; An
                     unidirectional cDNA library generated from [Root
                     tips(0.3cm-long) from  adult plants]. cDNA was prepared
                     from polyA+ mRNA using SuperScript  Plasmid System Kit
                     (Invitrogen). The double-strand cDNAs were  fractionated
                     in a sepharose CL-2B 40cm-columns and fragments sizing
                     between 0.8 and 1.5 Kb were directionally cloned into the
                     vector. Details  of each source of RNA and library
```

construction can be obtained at
http://sucest.lad.ic.unicamp.br/public"
ORIGIN

Query Match 7.8%; Score 35.4; DB 13; Length 664;
Best Local Similarity 46.2%; Pred. No. 44;
Matches 117; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy         22 GCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGC 81
              || | ||      ||| || |   |    || |   | ||  | | ||  |||  | | |
Db        521 GCCTGCGAGTATGTCGGAGCTGAGGCACCTGGTTGCGTTGCAGACCTTGTACATGGCGAC 462

Qy         82 GGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATT 141
              |   |||      |   |  |||||||  |      |||||  ||  |   |   |||||
Db        461 GTGCGCGCGGGGCGCCATCCCGGCGGCGACCCCCACGGCCTGGCCAAGAACCTGAGCGCC 402

Qy        142 TATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGAT 201
                ||  |||| ||||  |  |  |||    | | |||    ||  ||| |       ||
Db        401 CTGCACGACGGCTCCGGTGGCCGTGCTGGCGGTGTGCGTGCCGTGGCCGGCGTCATCGAC 342

Qy        202 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAAT 261
              | ||          || ||    |  ||     | ||||| |||   | |  | ||| |
Db        341 CGGCGACGAGGAGTTGCTGCCGGCGGTGGGGACGGACATGAAGGAGCGAGCTCCGATGAG 282

Qy        262 AGTACTATTGAAC 274
                | | |||  ||
Db        281 CTTGTTGTTGCAC 269
```

RESULT 4
BX251172
LOCUS BX251172 675 bp mRNA linear EST 24-FEB-2003
DEFINITION BX251172 Pinus pinaster differenciating xylem adult Pinus pinaster cDNA clone PP046C03, mRNA sequence.
ACCESSION BX251172
VERSION BX251172.1 GI:28513831
KEYWORDS EST.
SOURCE Pinus pinaster
ORGANISM Pinus pinaster
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 675)
AUTHORS Canton,F.R., Le Provost,G., Garcia,V., Barre,A., Frigerio,J.-M., Paiva,J., Fevereiro,P., Avila,C., Mouret,J.-F., Brach,J., de Daruvar,A., Canovas,F.M. and Plomion,C.
TITLE Transcriptome analysis of wood formation in maritime pine
JOURNAL Unpublished (2003)
COMMENT Contact: Frigerio JM
Genetique et Amelioration 69
INRA
route d'Arcachon 33612 Cestas CEDEX France
Email: Frigerio@pierroton.inra.fr
Email: Frigerio@pierroton.inra.fr.
FEATURES Location/Qualifiers
source 1. .675
/organism="Pinus pinaster"
/mol_type="mRNA"
/strain="ecotype: Corsican"
/db_xref="taxon:71647"
/clone="PP046C03"
/tissue_type="differenciating xylem"
/dev_stage="adult"
/clone_lib="Pinus pinaster differenciating xylem adult"
/note="Vector: Uni-Zap XR lambda (Stratagene); Site_1: Eco RI; Site_2: Xho I; A composite cDNA library was made with mRNA isolated from normal, compression, opposite, early and late wood of Maritime pine uni-directionally cloned into Uni-ZAP XR using the ZAP-cDNA Synthesis kit (Stratagene). pBluescript SK(-) plasmids were obtained by in vivo mass excision. The nucleotide sequence of the 5'end was obtained by automated sequencing with the T3 primer by GENOME EXPRESS, Meylan, France"
ORIGIN

Query Match 7.6%; Score 34.8; DB 13; Length 675;
Best Local Similarity 57.3%; Pred. No. 66;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        192 TAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGG 251
              ||||  |  |||| | | | |   | || | || ||||||   ||| |||||   | |||
Db        526 TAAAATTCCTCAGATAGATGCTATTATTTTGACACATGAACATGCAGATGCATTGCTGGG 585

Qy        252 TGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
              |  |||| ||  ||  | || ||   |  | ||   | ||   ||| ||
Db        586 TTTGGATGATGTTAGGAGTGTACAACCATATAACCCGTACAATGATATTG 635
```

RESULT 5
BI959328/c
LOCUS BI959328 699 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEn0019E09f Hordeum vulgare rachis EST library HVcDNA0015 (normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEn0019E09f, mRNA sequence.
ACCESSION BI959328
VERSION BI959328.1 GI:16310583
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 699)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex rachis cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 275
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence stop: 452.
FEATURES Location/Qualifiers
source 1. .699
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEn0019E09f"
/tissue_type="Rachis"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare rachis EST library HVcDNA0015 (normal)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and

sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 699;
Best Local Similarity 50.3%; Pred. No. 76;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy          37 GTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCAT 96
               ||| ||  ||||  ||||  | ||  ||||      || |||| |||| ||||      |
Db         202 GTACTTGGTGGCGTCGCTCCCGTTGAGGTCGACGTCGAAGGGGGAGGCGTCGGTGGCGTT 143

Qy          97 AACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCC 156
               || | |  ||||        ||||| |||  ||  |  | ||    | ||   ||
Db         142 AAGGTCCACGGCGCCGCCGACGGCCAGGAAGCAGAGAAGCGCCGCGAGCACGGCGAGGGA 83

Qy         157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGC 205
               |  | |  || | |||   ||   | | ||   ||   | ||  | ||
Db          82 GGGGAGGGAGCCCATGCCGCGCGCTGTGGTGTGTGTGTGTGTGTGCGGC 34
```

RESULT 6
AV058630
LOCUS AV058630 296 bp mRNA linear EST 23-JUN-1999
DEFINITION AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA clone 1810054H02, mRNA sequence.
ACCESSION AV058630
VERSION AV058630.1 GI:5158377
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 296)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"
/clone_lib="Mus musculus pancreas C57BL/6J adult"
ORIGIN

Query Match 7.5%; Score 34.2; DB 9; Length 296;
Best Local Similarity 53.3%; Pred. No. 68;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy         251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCG 310
               |||  ||  | |||  | |  |  | ||  ||  ||| |  |||||  | ||  | |  |
Db          13 GTGGAGACCACAGTTTTTTCTAGGTCACCAAGGCTGGGTAAAGAAAAGAAGCATCAACTG 72

Qy         311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
               |||||| | | | |||||    ||    | ||    || | | |  || | ||| |  |
Db          73 TCCAGTGAAGCACAAAAAATGAAGAAGAAAATGATTACCCAAAAGTTGGAATTAAGGTAG 132

Qy         371 CGCTGGTTAATCAGA 385
               | ||   ||  |||
Db         133 AGATGCCAAAATAGA 147
```

RESULT 7
AZ959284/c
LOCUS AZ959284 669 bp DNA linear GSS 27-APR-2001
DEFINITION 2M0226G20R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0226G20 R, genomic survey sequence.
ACCESSION AZ959284
VERSION AZ959284.1 GI:13830511
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 669)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0226 row: G column: 20
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 669.
FEATURES Location/Qualifiers
source 1. .669
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0226G20"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 7.5%; Score 34.2; DB 28; Length 669;
Best Local Similarity 54.3%; Pred. No. 97;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              | |||  ||   || ||  || |  || |  ||  || |  |  || |||| ||||||
Db        235 CTGGAGGCATGCTTAAGGTTTCAGAAGCAATGTAACGTTCCCATTGTGCTTTCTCTGCTT 176

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              | | |  || |||||||| |     | |  || || ||  |  | |  | |||| |||
Db        175 CCTGCTTCAGGTCAATATGAGCTCTCAGCTGTCCCTGCTGCCATGCCTTTACTGGCCCAT 116

Qy        241 GCAGGCC 247
              || |
Db        115 CATGGAC 109
```

RESULT 8
BZ569281
LOCUS BZ569281 978 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_8328.y2 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_8328, genomic survey sequence.
ACCESSION BZ569281
VERSION BZ569281.1 GI:27203648
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 978)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .978
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_8328"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."
ORIGIN

Query Match 7.5%; Score 34.2; DB 28; Length 978;
Best Local Similarity 53.3%; Pred. No. 1.1e+02;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        213 CCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGA 272
              | || | | |||||||   || | | |||  ||| ||   | ||| | || | ||||||
Db        519 CGGTATCGATACCCATTCGATTGAAAATGATGGCGAGTTCGTTGATCAGAGCAATATTGA 578

Qy        273 ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTC 332
              |  | | |  |  ||||    |   || || | ||  | | |  || |||     |
Db        579 GATCTCGCTGCGTATTTTCGATCACCTTGGCAGCCTCAGCGACTTTAATGCTGCTCGCCT 638

Qy        333 CGATATTACTGTCGG 347
              |  | ||| | | |
Db        639 TGTGAGTACCGGCAG 653
```

RESULT 9
BG671547
LOCUS BG671547 1036 bp mRNA linear EST 30-APR-2001
DEFINITION DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5', mRNA sequence.
ACCESSION BG671547
VERSION BG671547.1 GI:13893646
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1036)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
TITLE Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
MEDLINE 22056133
PUBMED 12060780
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR PRimers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1036
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRNBUB05"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/clone_lib="Rat DRG Library"
ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 1036;
Best Local Similarity 49.2%; Pred. No. 1.2e+02;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy         11 TAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTC 70
              | ||||   ||  | |  |||    |||  || | | || |||| |     ||||| |
Db        168 TGAAAGATACATGAGTGACAGTGTCCGTGATTGCAGACATTGCTGTATGATGCGTCCTGT 227

Qy         71 CACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAA 130
              |   | | | |  ||  ||||| ||  ||  | |   || | |   |  ||     |
Db        228 CTGTAGGAGTCCTCGCTGGTAACGATGCCGAGAGTGCACATCGATATCGGTGGGCGGTTA 287
```

```
Qy        131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
              | |       |  |  ||   | |    | ||| |||   ||| ||||  ||| || ||
Db        288 CATCTGAAGTGGAAGAGGTGGCTCTTTCGAACTCTGACTTTGATCTGTCAAGGGACGCTT 347

Qy        191 GTA 193
              |||
Db        348 GTA 350


RESULT 10
BZ565783/c
LOCUS       BZ565783               1137 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_5750.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_5750, genomic survey sequence.
ACCESSION   BZ565783
VERSION     BZ565783.1  GI:27194113
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1137)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_5750"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 28;  Length 1137;
  Best Local Similarity   53.3%;  Pred. No. 1.2e+02;
  Matches   72;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        213 CCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGA 272
              | || | | ||||||| || | | |||  ||| ||   |  ||| | || | ||||||
Db        272 CGGTATCGATACCCATTCGATTGAAAATGATGGCGAGTTCGTTGATCAGAGCAATATTGA 213

Qy        273 ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTC 332
              |  | | |  | ||||    |  || || | ||  | | |  || |||    |
Db        212 GATCTCGCTGCGTATTTTCGATCACCTTGGCAGCCTCAGCGACTTTAATGCTGCTCGCCT 153

Qy        333 CGATATTACTGTCGG 347
              |  | ||| | | |
Db        152 TGTGAGTACCGGCAG 138


RESULT 11
BB581987/c
LOCUS       BB581987               234 bp    mRNA    linear   EST 30-NOV-2000
DEFINITION  BB581987 RIKEN full-length enriched, adult male colon Mus musculus
            cDNA clone 9030005O05 5', mRNA sequence.
ACCESSION   BB581987
VERSION     BB581987.1  GI:11478531
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
            Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
            Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
            Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
            Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
            Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9030005O05"
                     /sex="male"
                     /tissue_type="colon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male colon"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 10;  Length 234;
```

```
  Best Local Similarity   51.3%;  Pred. No. 70;
  Matches   79;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         35 TCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATC 94
              | ||  ||  | |     | |   || ||  || |     ||    || | || || ||   |
Db        200 TAGTTTTTGGTAGGTCGGTTTATGCGGGTTTCTTCTTCCACCAGGCCCGCCGCCGTTTCC 141

Qy         95 ATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTT 154
                |||| |||  ||  ||   ||       |   |  || | ||||||| ||  ||  ||
Db        140 CGAACGCCGGGTGCTTTATAGCCTTGTTCAAAATATGGTAGCGCATTTACCACCACCTTT 81

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||| | ||   || | ||| |  ||  |||  ||
Db         80 CCGGTTACCAAGCCCGTGCCCGCCATTGCGCGGC 47


RESULT 12
CA711477/c
LOCUS       CA711477                500 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wdk2c.pk014.o3 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.o3 5'
            end, mRNA sequence.
ACCESSION   CA711477
VERSION     CA711477.1  GI:25433270
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone="wdk2c.pk014.o3"
                     /tissue_type="kernel"
                     /clone_lib="wdk2c"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
                     days after anthesis."
ORIGIN

  Query Match             7.5%;  Score 34;  DB 14;  Length 500;
  Best Local Similarity   54.5%;  Pred. No. 97;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  |  ||| |||||||   ||   |  | ||| ||       | | ||  | ||| |
Db        274 AGNTGNTGCAGTCGCAGCCNCCGGTCTCCCCGGCCGTNTCGAACCCGCCGGAGCTCCCCN 215

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGAC 126
                 ||| ||||| ||  || |  | |||||||||     |    ||||   ||
Db        214 TGTGGGAGGCGGTGGGCATGACCATGGCGGCGGCGCCGGCGGTGGCCNCCAC 163


RESULT 13
AZ180696
LOCUS       AZ180696                500 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_0181_A1_C06_SP6E Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library Strongylocentrotus purpuratus
            genomic clone Plate=181 Col=11 Row=E, genomic survey sequence.
ACCESSION   AZ180696
VERSION     AZ180696.1  GI:8353071
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
            Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H.
            and Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
   PUBMED   10920195
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 181  row: E  column: 11
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 500.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7668"
                     /clone="Plate=181 Col=11 Row=E"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 28;  Length 500;
  Best Local Similarity   50.3%;  Pred. No. 97;
  Matches   82;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        133 TTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGT 192
              ||   || | ||    | | |  || | | | ||    || |    |    ||||   |
Db          8 TTNTCCAATAATATCCATGATAACGATGATGATGATGATGATGATGATGATTGA 67

Qy        193 AAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGT 252
                | ||||| |    | |     || || |    |||||   |   |||  |   | | |
Db         68 GACGATGATGACGATGATGATGATGATGATGATGATGAATACGATGATGATGATGATGAT 127

Qy        253 GCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
              |  ||| ||| || || |||| | | | |  ||| || |||||
Db        128 GATGATGATAATAATAATGAAATTAATAATGATGATTACAGAA 170


RESULT 14
AA746477/c
LOCUS       AA746477                365 bp    mRNA    linear   EST 27-JAN-1998
DEFINITION  nw62e02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
            similar to TR:O00555 O00555 NEURONAL CALCIUM CHANNEL ALPHA 1A
            SUBUNIT ISOFORM A-1. ;contains element TAR1 repetitive element ;,
            mRNA sequence.
ACCESSION   AA746477
VERSION     AA746477.1  GI:2786463
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 365)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 661    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1251194"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 365;
  Best Local Similarity   56.9%;  Pred. No. 97;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  ||||||   ||||||   | ||   | ||| |||  | ||| |||   |  ||||
Db        119 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGTGGCCAGGCCGGCCGGGCGCCACC 60

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |  || |  ||| ||||  ||   | ||   | | |   | | ||||  |
Db         59 AGCGGCCCTCGGAGGTACCCAGGCCCGCACGGCCGACGCTACTGGCCGG 11


RESULT 15
CD374421/c
LOCUS       CD374421                 558 bp    mRNA    linear   EST 19-AUG-2003
DEFINITION  TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)
            Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION   CD374421
VERSION     CD374421.1  GI:31229939
KEYWORDS    EST.
SOURCE      Wuchereria bancrofti
  ORGANISM  Wuchereria bancrofti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Wuchereria.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
  TITLE     Toward the characterization of the Wuchereria bancrofti
            microfilarial transcriptome with comparisons to those of Brugia
            malayi and Onchocerca volvulus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Thomas B. Nutman
            Laboratory of Parasitic Diseases
            NIH
            Building 4 Room 126, Bethesda, MD 20892-0425, USA
            Email: tnutman@niaid.nih.gov
            Seq primer: pBluescript T3.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Wuchereria bancrofti"
                     /mol_type="mRNA"
                     /db_xref="taxon:6293"
                     /clone="TNWbmfC2H3"
                     /dev_stage="microfilaria"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="Wuchereria bancrofti microfilaria cDNA
                     (SAW95SjL-WbMf)"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 85,000 microfilariae
                     isolated from the blood of an infected individual from
                     Guyana, South America and converted to double-stranded
                     cDNA using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNA pol I. The library has 2.9 x 10E5
                     independent recombinants and the average insert size is
                     ~1kb. The library was constructed by Sandra J. Laney. The
                     library is available from Dr.S.A.Williams, email:
                     genome@smith.edu."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 558;
  Best Local Similarity   60.7%;  Pred. No. 1.2e+02;
  Matches   65;  Conservative    0;  Mismatches   40;  Indels    2;  Gaps    1;

Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              ||| |||| |||||||| |     || ||||| |||| |    | ||  | |
Db        161 TGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCCCAAATTCTANGCT 102

Qy        136 AGCATTTATC--AGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||| |  ||  ||||  ||  | | |||| ||||||||| ||
Db        101 GACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55


Search completed: March 18, 2004, 20:11:33
Job time : 2035.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	2067	1	SEU43280	U43280 Salmonella
2	452.8	99.3	5103	1	STAJ2301	AJ002301 Salmonell
3	452.8	99.3	22411	1	AE008749	AE008749 Salmonell
4	451.2	98.9	254050	1	AL627269	AL627269 Salmonell
c 5	451.2	98.9	301983	1	AE016840	AE016840 Salmonell
6	382.8	83.9	1048	1	STAGFBA	AJ000514 Salmonell
7	348.8	76.5	2889	1	CSP515700	AJ515700 Citrobact
8	334.4	73.3	10370	1	AE015131	AE015131 Shigella
9	332.8	73.0	4680	1	ECCSGABDG	X90754 E.coli csgG
10	332.8	73.0	10346	1	AE000205	AE000205 Escherich
11	332.8	73.0	15047	1	D90741	D90741 Escherichia
12	332.8	73.0	292504	1	AE016981	AE016981 Shigella
13	332.8	73.0	306358	1	AE016759	AE016759 Escherich
14	331.2	72.6	1711	1	AF275733	AF275733 Escherich
15	331.2	72.6	10190	1	AE005315	AE005315 Escherich
16	331.2	72.6	327773	1	AP002554	AP002554 Escherich
17	313.4	68.7	2920	1	CFR515701	AJ515701 Citrobact
18	271	59.4	2883	1	ESA515702	AJ515702 Enterobac
19	184	40.4	1212	1	ECO131756	AJ131756 Escherich
c 20	47.8	10.5	10709	1	AE015532	AE015532 Shewanell
21	47.4	10.4	301214	1	AE016786	AE016786 Pseudomon
22	44.2	9.7	11160	1	AE008209	AE008209 Agrobacte
c 23	44.2	9.7	11476	1	AE009405	AE009405 Agrobacte
24	41.4	9.1	42602	2	AC116971	AC116971 Trypanoso
c 25	41	9.0	7218	6	I66494	I66494 Sequence 14
c 26	40.6	8.9	62381	5	BX640594	BX640594 Zebrafish
27	40.4	8.9	168457	2	AC111425	AC111425 Rattus no
c 28	40.4	8.9	178934	2	AC127966	AC127966 Rattus no
29	40.2	8.8	193553	10	AC124196	AC124196 Mus muscu
c 30	39.8	8.7	295500	1	AP005954	AP005954 Bradyrhiz
c 31	38.2	8.4	298406	3	CEY75B8A	AL033514 Caenorhab
c 32	37.8	8.3	174732	2	AL136161	AL136161 Homo sapi
c 33	37.8	8.3	191567	9	AL138702	AL138702 Human DNA
c 34	37.6	8.2	2000	6	AX655393	AX655393 Sequence
35	37.4	8.2	120937	10	AL591911	AL591911 Mouse DNA
c 36	37.4	8.2	225016	2	AC063967	AC063967 Mus muscu
37	37.2	8.2	42310	3	CBRG01B5	AC084425 Caenorhab
c 38	37	8.1	164504	2	AC123240	AC123240 Rattus no
39	36.8	8.1	252420	3	AE014841	AE014841 Plasmodiu
c 40	36.6	8.0	92378	5	BX537249	BX537249 Zebrafish
c 41	36.6	8.0	110000	2	AC095863_06	Continuation (7 of
42	36.6	8.0	196383	2	AC113159	AC113159 Rattus no
43	36.6	8.0	230058	2	AC108229	AC108229 Rattus no
44	36.6	8.0	254497	2	AC119520	AC119520 Rattus no
45	36.6	8.0	323930	1	AP003194	AP003194 Clostridi

ALIGNMENTS

RESULT 1
SEU43280
LOCUS SEU43280 2067 bp DNA linear BCT 14-FEB-1996
DEFINITION Salmonella enteritidis agfBAC operon: fimbrin-like protein precursor (agfB), thin aggregative fimbriae precursor (agfA), and AgfC (agfC) genes, complete cds.
ACCESSION U43280
VERSION U43280.1 GI:1184712
KEYWORDS .
SOURCE Salmonella enteritidis
ORGANISM Salmonella enteritidis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 2067)
AUTHORS Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,

```
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match           100.0%;  Score 456;  DB 1;  Length 2067;
  Best Local Similarity 100.0%;  Pred. No. 3.8e-123;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        696 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 755

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 815

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        816 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 875

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        876 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 935

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        936 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 995

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        996 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 1055

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1056 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 1115

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1116 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 1151


RESULT 2
STAJ2301
LOCUS       STAJ2301                5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATPAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
     gene            3344. .3376
                     /gene="csgBA"
     -35_signal      3344. .3349
                     /gene="csgBA"
     -10_signal      3371. .3376
                     /gene="csgBA"
     gene            3469. .3924
                     /gene="csgB"
     CDS             3469. .3924
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="CAA05316.1"
                     /db_xref="GI:2739238"
                     /db_xref="GOA:P55226"
                     /db_xref="SWISS-PROT:P55226"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     3469. .3531
                     /gene="csgB"
     mat_peptide     3532. .3921
                     /gene="csgB"
                     /product="CsgB protein"
                     /function="nucleator"
     gene            3966. .4421
                     /gene="csgA"
     CDS             3966. .4421
                     /gene="csgA"
                     /function="fiber subunit"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="CAA05317.1"
                     /db_xref="GI:2739239"
                     /db_xref="GOA:P55225"
                     /db_xref="SWISS-PROT:P55225"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     3966. .4025
                     /gene="csgA"
     mat_peptide     4026. .4418
                     /gene="csgA"
                     /product="CsgA protein"
                     /function="fiber subunit"
     gene            4483. .4809
                     /gene="csgC"
     CDS             4483. .4809
                     /gene="csgC"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgC protein"
                     /protein_id="CAA05318.1"
```

```
                     /db_xref="GI:2739240"
                     /db_xref="SWISS-PROT:P55227"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
     sig_peptide     4483. .4533
                     /gene="csgC"
     mat_peptide     4534. .4806
                     /gene="csgC"
                     /product="CsgC"
     gene            4940. .5103
                     /gene="ORF103"
     CDS             4940. .>5103
                     /gene="ORF103"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA05319.1"
                     /db_xref="GI:2739241"
                     /db_xref="SWISS-PROT:O54295"
                     /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                     CDISTHAENI"
ORIGIN

  Query Match             99.3%;  Score 452.8;  DB 1;  Length 5103;
  Best Local Similarity   99.6%;  Pred. No. 3.4e-122;
  Matches  454;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3469 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 3528

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       3529 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 3588

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3589 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 3648

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3649 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 3708

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3709 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 3768

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       3769 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 3828

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3829 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 3888

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       3889 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 3924


RESULT 3
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .22411
                     /organism="Salmonella typhimurium LT2"
                     /mol_type="genomic DNA"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            434. .1308
                     /gene="phoH"
                     /note="synonym: STM1126"
     RBS             434. .439
                     /gene="phoH"
                     /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS             454. .1308
                     /gene="phoH"
                     /note="similar to E. coli PhoB-dependent, ATP-binding pho
                     regulon component; may be helicase; induced by P
                     starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                     aa), 92% identity in aa 71 - 354"
                     /codon_start=1
                     /transl_table=11
                     /product="PhoB-dependent ATP-binding pho regulon
                     component"
                     /protein_id="AAL20057.1"
                     /db_xref="GI:16419642"
                     /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                     RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                     GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                     YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                     LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                     QRTLHAYS"
     gene            complement(1414. .2303)
```

```
                /gene="STM1127"
CDS             complement(1414. .2295)
                /gene="STM1127"
                /note="similar to E. coli orf, hypothetical protein
                (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                identity in aa 15 - 264"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /protein_id="AAL20058.1"
                /db_xref="GI:16419643"
                /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                QARESKLRTMNALDMTK"
RBS             complement(2298. .2303)
                /gene="STM1127"
                /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
gene            complement(2580. .4088)
                /gene="STM1128"
CDS             complement(2580. .4076)
                /gene="STM1128"
                /note="similar to E. coli putative cotransporter
                (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                identity in aa 7 - 478"
                /codon_start=1
                /transl_table=11
                /product="putative sodium/glucose cotransporter"
                /protein_id="AAL20059.1"
                /db_xref="GI:16419644"
                /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
RBS             complement(4083. .4088)
                /gene="STM1128"
                /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
gene            complement(4413. .5099)
                /gene="STM1129"
CDS             complement(4413. .5093)
                /gene="STM1129"
                /note="similar to E. coli putative enzyme (AAC76255.1);
                Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                226"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL20060.1"
                /db_xref="GI:16419645"
                /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                WFSHAVKR"
RBS             complement(5094. .5099)
                /gene="STM1129"
                /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
gene            5599. .6759
                /gene="STM1130"
CDS             5599. .6759
                /gene="STM1130"
                /note="similar to E. coli orf, hypothetical protein
                (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                identity in aa 40 - 403"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL20061.1"
                /db_xref="GI:16419646"
                /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
gene            6793. .7497
                /gene="STM1131"
RBS             6793. .6798
                /gene="STM1131"
                /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
CDS             6805. .7497
                /gene="STM1131"
                /note="similar to E. coli orf, hypothetical protein
                (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                identity in aa 4 - 239"
                /codon_start=1
                /transl_table=11
                /product="putative outer membrane protein"
                /protein_id="AAL20062.1"
                /db_xref="GI:16419647"
                /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                QTRYRVGLQYHF"
gene            7768. .9060
                /gene="STM1132"
RBS             7768. .7773
                /gene="STM1132"
                /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
CDS             7780. .9060
                /gene="STM1132"
                /note="similar to E. coli putative transport protein
                (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                identity in aa 25 - 418"
                /codon_start=1
                /transl_table=11
                /product="putative sugar transport protein"

  Query Match             99.3%;  Score 452.8;  DB 1;  Length 22411;
  Best Local Similarity   99.6%;  Pred. No. 3.6e-122;
  Matches  454;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17272 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 17331

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      17332 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 17391

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17392 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 17451

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17452 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 17511

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17512 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 17571

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      17572 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 17631

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17632 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 17691

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
                ||||||||||||||||||||||||||||||||||||
Db      17692 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 17727


RESULT 4
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAARAKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap

  Query Match           98.9%;  Score 451.2;  DB 1;  Length 254050;
  Best Local Similarity  99.3%;  Pred. No. 1.2e-121;
  Matches  453;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88409 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 88468

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      88469 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 88528

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88529 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 88588

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      88589 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGAAAATAATCGGGCG 88648

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88649 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 88708

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      88709 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 88768

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88769 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 88828

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      88829 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 88864


RESULT 5
AE016840/c
LOCUS       AE016840              301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
```

/transl_table=11
/product="transposase for insertion sequence element IS200"
/protein_id="AAO69359.1"
/db_xref="GI:29137798"
/translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"

gene complement(844. .1797)
/gene="flgL"
/locus_tag="t1736"

CDS complement(844. .1797)
/gene="flgL"
/locus_tag="t1736"
/note="corresponds to STY1223 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook-associated protein 3"
/protein_id="AAO69360.1"
/db_xref="GI:29137799"
/translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"

gene complement(1812. .3473)
/gene="flgK"
/locus_tag="t1737"

CDS complement(1812. .3473)
/gene="flgK"
/locus_tag="t1737"
/note="corresponds to STY1222 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook-associated protein 1"
/protein_id="AAO69361.1"
/db_xref="GI:29137800"
/translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"

gene complement(3538. .4488)
/gene="flgJ"
/locus_tag="t1738"

CDS complement(3538. .4488)
/gene="flgJ"
/locus_tag="t1738"
/note="corresponds to STY1221 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar protein FlgJ"
/protein_id="AAO69362.1"
/db_xref="GI:29137801"
/translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"

gene complement(4488. .5591)
/gene="flgI"
/locus_tag="t1739"

CDS complement(4488. .5591)
/gene="flgI"
/locus_tag="t1739"
/note="corresponds to STY1220 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar P-ring protein precursor"
/protein_id="AAO69363.1"
/db_xref="GI:29137802"
/translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV RALNALGATPMDLMSILQSMQSAGCLRAKLEII"

gene complement(5597. .6295)
/gene="flgH"
/locus_tag="t1740"

CDS complement(5597. .6295)
/gene="flgH"
/locus_tag="t1740"
/note="corresponds to STY1219 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AAO69364.1"
/db_xref="GI:29137803"
/translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN MGWLQRFFLNLSPM"

gene complement(6350. .7132)
/gene="flgG"
/locus_tag="t1741"

CDS complement(6350. .7132)
/gene="flgG"
/locus_tag="t1741"
/note="distal rod protein; corresponds to STY1218 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein FlgG"
/protein_id="AAO69365.1"
/db_xref="GI:29137804"
/translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"

gene complement(7146. .7901)
/gene="flgF"
/locus_tag="t1742"

CDS complement(7146. .7901)
/gene="flgF"
/locus_tag="t1742"
/note="corresponds to STY1217 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein FlgF"
/protein_id="AAO69366.1"
/db_xref="GI:29137805"
/translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT DMIANARRFEMQMKVITSVDENEGRANQLLSMS"

gene complement(7922. .9133)
/gene="flgE"
/locus_tag="t1743"

```
     CDS             complement(7922. .9133)
                     /gene="flgE"
                     /locus_tag="t1743"
                     /note="corresponds to STY1216 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook protein FlgE"
                     /protein_id="AAO69367.1"
                     /db_xref="GI:29137806"
                     /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                     GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

  Query Match             98.9%;  Score 451.2;  DB 1;  Length 301983;
  Best Local Similarity   99.3%;  Pred. No. 1.2e-121;
  Matches  453;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37807 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 37748

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      37747 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 37688

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37687 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 37628

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      37627 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGAAAATAATCGGGCG 37568

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37567 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 37508

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      37507 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 37448

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37447 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 37388

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      37387 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 37352


RESULT 6
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             83.9%;  Score 382.8;  DB 1;  Length 1048;
  Best Local Similarity   94.6%;  Pred. No. 1.3e-101;
  Matches  435;  Conservative    0;  Mismatches    7;  Indels   18;  Gaps    3;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||||||||||||||||| |||   ||||||||||
Db         14 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGAGCC---GATTGCAACC 70

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         71 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 130

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 190

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGA---AGGAGGAAATAATCGG 237
              ||||| |||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        191 CGCCATGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAACGAGGAGGAAATAATCGG 250

Qy        238 GCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCC 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 GCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCC 310

Qy        298 AACGATGCCAGTATATCGCAAAGCGCTTACGGTAA----------TAGTGCAGCTATT 345
              |||||||||||||||||||||||||||||||||||          |||||| ||||||
Db        311 AACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTAGTGCGGCTATT 370

Qy        346 ATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCA 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 ATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCA 430

Qy        406 GTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCA 445
              |||||||||||||||||||||||||||||||| | ||
Db        431 GTTGTAGTGCAGAAACAGTCGCATATGGCTATTCAGGCCA 470
```

```
RESULT 7
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
     gene            HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
     CDS             KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
                     1623. .2078
                     /gene="csgB"
                     1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
     gene            /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
     CDS             QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
                     2119. .2571
                     /gene="csgA"
                     2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
ORIGIN               /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
```

```
Query Match             76.5%;  Score 348.8;  DB 1;  Length 2889;
Best Local Similarity   85.3%;  Pred. No. 1.4e-91;
Matches  389;  Conservative    0;  Mismatches   67;  Indels     0;  Gaps     0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1623 ATGAAAAACAAATTGTTATTTATGATGTTAACAGTACTGGGTGCGCCTGGAATTGCTTCC 1682

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||| |||||||| |||| |||||||||||||||||||||||||||||||||||||||||
Db       1683 GCAACAAGTTATGATTTAGCGCATTCCGAATATAACTTTGCGGTAAATGAATTAAGTAAG 1742

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1743 TCTTCATTTAATCAGGCAGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 1802

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1803 CGCCAGGAGGGCTCTAAACTTTTGTCGGTAGTTTCACAAGAAGGTGGAAGTAACCGGGCC 1862

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1863 AAAGTTGACCAATCAGGAGCTTATAATTTTGCGTATATTGCTCAGTCGGGTCATTCCAAT 1922

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1923 GATGCCAGTATATCGCAAAGTAATTACGGTAATACTGCGATGATTATCCAGAAAGGTTCT 1982

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1983 GGAAATAAAGCAAATATTACTCAGTATGGTACGCAAAAAACAGCAGTTGTAGTGCAGAGA 2042

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||||||||||||||||||||||||||||||
Db       2043 CAGTCGCAAATGGCAATTCGCGTTATTCAACGCTAA 2078


RESULT 8
AE015131
LOCUS
DEFINITION  AE015131                10370 bp    DNA     linear   BCT 18-OCT-2002
            Shigella flexneri 2a str. 301 section 94 of 412 of the complete
ACCESSION   genome.
VERSION     AE015131 AE005674
KEYWORDS    AE015131.1  GI:24051313
SOURCE      .
  ORGANISM  Shigella flexneri 2a str. 301
            Shigella flexneri 2a str. 301
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 10370)
  AUTHORS   Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
            Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
            Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
            Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
            Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
  TITLE     Genome sequence of Shigella flexneri 2a: insights into
            pathogenicity through comparison with genomes of Escherichia coli
            K12 and O157
  JOURNAL   Nucleic Acids Res. 30 (20), 4432-4441 (2002)
  PUBMED    12384590
REFERENCE   2  (bases 1 to 10370)
  AUTHORS   Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B.,
            Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y.,
            Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y.,
            Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
            of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P R
            China
FEATURES             Location/Qualifiers
     source          1. .10370
                     /organism="Shigella flexneri 2a str  301"
```

/mol_type="genomic DNA"
/strain="301"
/serotype="2a"
/db_xref="taxon:198214"
tRNA complement(127. .214)
/product="tRNA-Ser"
/note="anticodon: GGA"
gene 410. .1387
/gene="ycdW"
/locus_tag="SF1027"
/pseudo
CDS 410. .1387
/gene="ycdW"
/locus_tag="SF1027"
/note="similar to Escherichia coli K12 putative dehydrogenase gi: 1787270 (326 aa). BLAST with identity of 99% in 326 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11
gene 1442. .2179
/gene="ycdX"
/locus_tag="SF1028"
CDS 1442. .2179
/gene="ycdX"
/locus_tag="SF1028"
/note="Residues 1 to 245 of 245 are 99 pct identical to residues 1 to 245 of a 245 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287168.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42652.1"
/db_xref="GI:24051314"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE PVFAPHDKATNTQAMISTIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS SFLHSRKGSEDNCRAVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPLER ILNVSPRRLLNFLESRGMAPIAEFADL"
gene 2203. .2757
/gene="ycdY"
/locus_tag="SF1029"
CDS 2203. .2757
/gene="ycdY"
/locus_tag="SF1029"
/note="Residues 1 to 184 of 184 are 99 pct identical to residues 1 to 184 of a 184 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287169.1"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase component"
/protein_id="AAN42653.1"
/db_xref="GI:24051315"
/translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEDATEAEVRAFLSERGMLL ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW RTMAPLTRDAISAMWDELEEDSEE"
gene 2811. .3350
/gene="ycdZ"
/locus_tag="SF1030"
CDS 2811. .3350
/gene="ycdZ"
/locus_tag="SF1030"
/note="Residues 1 to 179 of 179 are 98 pct identical to residues 1 to 179 of a 179 aa protein from Escherichia coli K12 ref: NP_415554.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42654.1"
/db_xref="GI:24051316"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLVFGYAMKNSGLW LAARSAKTAHREQEIKNKA"
gene complement(3414. .4189)
/gene="csgG"
/locus_tag="SF1031"
/pseudo
CDS complement(3414. .4189)
/gene="csgG"
/locus_tag="SF1031"
/note="similar to Escherichia coli K12 curli production assembly-transport component, 2nd curli operon gi: 1787274 (278 aa). BLAST with identity of 97% in 257 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11
gene complement(4454. .5662)
/locus_tag="SF1032"
CDS complement(4454. .5662)
/locus_tag="SF1032"
/note="Residues 1 to 402 of 402 are 100 pct identical to residues 1 to 402 of a 402 aa protein from Plasmid R100 ref: NP_052925.1"
/codon_start=1
/transl_table=11
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translation="MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTLTLTEL GRNLPTKARTKHNIKRIDRLLGNRHLHKERLAVYRWHASFICSGNTMPIVLVDWSDIR EQKRLMVLRASVALHGRSVTLYEKAFPLSEQCSKKAHDQFLADLASILPSNTTPLIVS DAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAENWKPISNLHDMSSSHSKTLGYKR LTKSNPISCQILLYKSRSKGRKNQRSTRTHCHHPSPKIYSASAKEPWILATNLPVEIR TPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWL AGVHAQKQGWDKHFQANTVRNRNVLSTVRLGMEVLRHSGYTITREDSLVAATLLTQNL FTHGYVLGKL"
gene complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"
CDS complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"
/note="Residues 1 to 129 of 129 are 98 pct identical to residues 1 to 129 of a 129 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287173.1"
/codon_start=1
/transl_table=11
/product="curli production assembly-transport component, 2nd curli operon"
/protein_id="AAN42656.1"
/db_xref="GI:24051318"
/translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDSEKTVVFAL IQTEEALNRRQINQALLSTDDLAHDEF"
gene complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"
CDS complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"
/note="Residues 11 to 226 of 226 are 100 pct identical to residues 1 to 216 of a 216 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=11
/product="putative 2-component transcriptional regulator for 2nd curli operon"
/protein_id="AAN42657.1"
/db_xref="GI:24051319"
/translation="MRSIKKWGFIMFNEVHSIHGHTLLLITKPSLQATALLQHLKQSL

```
                     AITGKLHNIQRSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPED
                     YPYRDIENWPHINGVFYAMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNS
                     TESALLTHREKEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVS
                     WANDNLRR"
     gene            7766. .8248
                     /gene="csgB"
                     /locus_tag="SF1035"
     CDS             7766. .8248
                     /gene="csgB"
                     /locus_tag="SF1035"
                     /note="Residues 10 to 160 of 160 are 99 pct identical to
                     residues 1 to 151 of a 151 aa protein from Escherichia
                     coli O157:H7 EDL933 ref: NP_287175.1"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAN42658.1"
                     /db_xref="GI:24051320"
                     /translation="MYDQVQGDNMKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAV
                     NELSKSSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYI
                     DQAGSANDASISQGAYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR

  Query Match          73.3%;  Score 334.4;  DB 1;  Length 10370;
  Best Local Similarity  83.3%;  Pred. No. 2.6e-87;
  Matches  380;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db       7793 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 7852

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db       7853 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 7912

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db       7913 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 7972

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  ||||||||
Db       7973 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAATCGGGCA 8032

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db       8033 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 8092

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db       8093 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 8152

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||||||||||||||||| |
Db       8153 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAGTTGTAGTGCAGAGA 8212

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| ||||| |||||||| || ||||| |||
Db       8213 CAGTCGCAAATGGCAATTCGCGTGACACAACGTTAA 8248


RESULT 9
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
```

```
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           73.0%;  Score 332.8;  DB 1;  Length 4680;
  Best Local Similarity  83.1%;  Pred. No. 7.5e-87;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||
Db       3233 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 3292

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db       3293 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 3352

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db       3353 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 3412

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db       3413 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 3472

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db       3473 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 3532

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db       3533 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 3592

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       3593 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 3652

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db       3653 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 3688


RESULT 10
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
```

```
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFPKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

  Query Match           73.0%;  Score 332.8;  DB 1;  Length 10346;
  Best Local Similarity   83.1%;  Pred. No. 7.7e-87;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db       8497 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 8556

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db       8557 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 8616

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db       8617 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 8676

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db       8677 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 8736

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | ||||||  ||||  |||||||| |||| |||||||| ||  ||||||  |||||||
Db       8737 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 8796

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| |||||||  |||     ||||||||||||||||||
Db       8797 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 8856

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       8857 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 8916

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db       8917 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 8952


RESULT 11
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
            .The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                    Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .15047
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:83333"
                     /map="23.7-24.0 min"
                     /clone="Kohara clone #231"
                     /note="Nucleotide position 1097101-1112147 from the
                     initiation site of ThrA (0 min.).~This clone is from
                     Kohara lambda miniset library."
     CDS             <1. .276
                     /note="unnamed protein product; ORF_ID:o230#5"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35822.1"
                     /db_xref="GI:4062605"
                     /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
                     WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     CDS             1152. .2090
                     /note="ORF_ID:o231#1
                     similar to SwissProt Accession Number P45637"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="Hypothetical 33.0 kd protein in proB-proA
                     intergenic region."
                     /protein_id="BAA35823.1"
                     /db_xref="GI:4062606"
                     /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
                     WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
                     AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
                     PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
                     GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
                     AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS                  2145. .2882
                     /note="unnamed protein product; ORF_ID:o231#2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35824.1"
                     /db_xref="GI:4062607"
                     /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                     EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                     PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                     SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
                     ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS                  2906. .3460
                     /note="ORF_ID:o231#3
                     similar to PIR Accession Number F64035"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein HI1543"
                     /protein_id="BAA35825.1"
                     /db_xref="GI:4062608"
                     /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                     DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                     ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                     RTMAPLTRDAISAMWDELEEDSEE"
CDS                  3514. .4053
                     /note="unnamed protein product; ORF_ID:o231#4"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35826.1"
                     /db_xref="GI:4062609"
                     /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                     SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                     IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                     LAARSAKTAHREQEIKNKA"
gene                 complement(4117. .4950)
                     /gene="csgG"
CDS                  complement(4117. .4950)
                     /gene="csgG"
                     /note="ORF_ID:o231#5
                     similar to SwissProt Accession Number P52103"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgG precursor."
                     /protein_id="BAA35827.1"
                     /db_xref="GI:4062610"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
gene                 complement(4977. .5393)
                     /gene="csgF"
CDS                  complement(4977. .5393)
                     /gene="csgF"
                     /note="ORF_ID:o231#6
                     similar to SwissProt Accession Number P52104"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgF precursor."
                     /protein_id="BAA35828.1"
                     /db_xref="GI:4062611"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene                 complement(5418. .5807)
                     /gene="csgE"
CDS                  complement(5418. .5807)
                     /gene="csgE"
                     /note="ORF_ID:o231#7
                     similar to SwissProt Accession Number P52105"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgE precursor."
                     /protein_id="BAA35829.1"
                     /db_xref="GI:4062612"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
gene                 complement(5812. .6462)
                     /gene="csgD"
CDS                  complement(5812. .6462)
                     /gene="csgD"
                     /note="ORF_ID:o231#8
                     similar to SwissProt Accession Number P52106"
                     /codon_start=1
                     /transl_table=11
                     /product="Probable csgAB operon transcriptional regulatory
                     protein."
                     /protein_id="BAA35830.1"
                     /db_xref="GI:4062613"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene                 7217. .7672
                     /gene="csgB"
CDS                  7217. .7672
```

```
Query Match           73.0%;  Score 332.8;  DB 1;  Length 15047;
Best Local Similarity 83.1%;  Pred. No. 7.8e-87;
Matches 379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||
Db       7217 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 7276

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || ||||||||||||||| || |||
Db       7277 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 7336

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| |||||||||||  || ||  ||||||||||     ||
Db       7337 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 7396

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | | || |||||
Db       7397 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 7456

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db       7457 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 7516

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||     ||||||||||||||||||
Db       7517 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 7576

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       7577 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 7636

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
```

```
Db       7637 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 7672


RESULT 12
AE016981
LOCUS       AE016981               292504 bp    DNA     linear   BCT 22-APR-2003
DEFINITION  Shigella flexneri 2a str. 2457T section 4 of 16 of the complete
            genome.
ACCESSION   AE016981 AE014073
VERSION     AE016981.1  GI:30040616
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 2457T
  ORGANISM  Shigella flexneri 2a str. 2457T
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Complete Genome Sequence and Comparative Genomics of Shigella
            flexneri Serotype 2a Strain 2457T
  JOURNAL   Infect. Immun. 71 (5), 2775-2786 (2003)
   PUBMED   12704152
REFERENCE   2  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .292504
                     /organism="Shigella flexneri 2a str. 2457T"
                     /mol_type="genomic DNA"
                     /strain="2457T"
                     /serotype="2a"
                     /db_xref="taxon:198215"
     gene            complement(223. .441)
                     /gene="infA"
                     /locus_tag="S0884"
     CDS             complement(223. .441)
                     /gene="infA"
                     /locus_tag="S0884"
                     /function="factor; Proteins - translation and
                     modification"
                     /note="residues 1 to 72 of 72 are 100.00 pct identical to
                     residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
                     /codon_start=1
                     /transl_table=11
                     /product="protein chain initiation factor IF-1"
                     /protein_id="AAP16348.1"
                     /db_xref="GI:30040617"
                     /translation="MAKEDNIEMQGTVLETLPNTMFRVELENGHVVTAHISGKMRKNY
                     IRILTGDKVTVELTPYDLSKGRIVFRSR"
     gene            complement(726. .1430)
                     /gene="aat"
                     /locus_tag="S0885"
     CDS             complement(726. .1430)
                     /gene="aat"
                     /locus_tag="S0885"
                     /function="enzyme; Aminoacyl tRNA synthetases, tRNA
                     modification"
                     /note="residues 1 to 234 of 234 are 100.00 pct identical
                     to residues 1 to 234 of 234 from Escherichia coli K-12 :
                     B0885"
                     /codon_start=1
                     /transl_table=11
                     /product="leucyl, phenylalanyl-tRNA-protein transferase"
                     /protein_id="AAP16349.1"
                     /db_xref="GI:30040618"
                     /translation="MRLVQLSRHSIAFPSPEGALREPNGLLALGGDLSPARLLMAYQR
                     GIFPWFSPGDPILWWSPDPRAVLWPESLHISRSMKRFHKRSPYRVTMNYAFGQVIEGC
                     ASDREEGTWITRGVVEAYHRLHELGHAHSIEVWREDELVGGMYGVAQGTLFCGESMFS
                     RMENASKTALLVFCEEFIGHGGKLIDCQVLNDHTASLGACEIPRRDYLNYLNQMRLGR
                     LPNNFWVPRCLFSPQE"
     gene            complement(1472. .3193)
                     /gene="cydC"
                     /locus_tag="S0886"
     CDS             complement(1472. .3193)
                     /gene="cydC"
                     /locus_tag="S0886"
                     /function="transport; Transport of small molecules: Other"
                     /note="residues 1 to 573 of 573 are 97.03 pct identical to
                     residues 1 to 573 of 573 from Escherichia coli K-12 :
                     B0886"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-binding component of cytochrome-related
                     transport"
                     /protein_id="AAP16350.1"
                     /db_xref="GI:30040619"
                     /translation="MRALLPYLALYKRHKWMLSLGIVLAIVTLLASIGLLTLSGWFLS
                     ASAVAGVAGLYSFNYMLPAAGVRGAAITRTAGRYFERLVSHDATFRVLQHLRIYTFSK
                     LLPLSPAGLARYRQGELLNRVVADVDTLDHLYLRVISPLVGAFVVIMVVTIGLSFLDF
                     TLAFTLGSVMLLTLFLMPPLFYRAGKSTGQNLTHLRGQYRQQLTAWLQGQAELTIFGA
                     SDRYRTQLENTEIQWLEAQRRQSELTALSQAIMLLIGALAVILMLWMASGGVGGNAQP
                     GALIALFVFCALAAFEALAPVTGAFQHLGQVIASAVRISDLTDQKPEVTFPDTQTRVA
                     DRVSLTLRDVQFTYPEQSQQALKGISLQVNAGEHIAILGRTGCGKSTLLQLLTRAWDP
                     QQGEILLNDSPIASLNEAALRQTISVVPQRVHLFSATLRDNLLLASPGSSDEALSEIL
                     RRVGLEKLLEDAGLNSWLGEGGRQLSGGELRRLAIARTLLHDAPLVLLDEPTEGLDAT
                     TESQILELLAKMMREKTVLMVTHRLRGLSRFQQIIVMDNGQIIEQGTHAELLARQGRY
                     YQFKQGL"
     gene            complement(3194. .4960)
                     /gene="cydD"
                     /locus_tag="S0887"
     CDS             complement(3194. .4960)
                     /gene="cydD"
                     /locus_tag="S0887"
                     /function="transport; Not classified"
                     /note="residues 1 to 588 of 588 are 92.85 pct identical to
                     residues 1 to 588 of 588 from Escherichia coli K-12 :
                     B0887"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-binding component of cytochrome-related
                     transport"
                     /protein_id="AAP16351.1"
                     /db_xref="GI:30040620"
                     /translation="MNKSRQKELTRWLKQQSVISQRWLNISRLLGFVSGILIIAQAWF
                     MARILQHMIMENIPREALLLPFTLLVLTFVLRAWVVWLRERVGYHAGQHIRFAIRRQV
                     LDRLQQAGPAWIQGKPAGSWATLVLEQIDDMHDYYARYLPQMALAVSVPLLIVVAIFP
                     SNWAAALILLGTAPLIPLFMALVGMGAADANRRNFLALARLSGHFLDRLRGMETLRIF
                     GRGEAEIESIRSASEDFRQRTMEVLRLAFLSSGILEFFTSLSIALVAVYFGFSYLGEL
                     DFGHYDTGVTLAAGFLALILAPEFFQPLRDLGTFYHAKAQAVGAADSLKTFMETPLAH
                     PQRGEAELALTDPLTIEAEDLFITSPEGKTLAGPLNFTLPAGQRAVLVGRSGSGKSSL
                     LNALSGFLSYQGSLRINGIELRDLSPESWRKHLSWVGQNPQLPAAKLRDNVLLARPDA
                     SEQELQAALDNAWVSEFLPLLPQGVDTPVGDQAARLSVGQAQRVAVARALLNPCSLLL
                     LDEPAASLDAHSEQRVMEALNAASLRQTTLMVTHQLEDLADWDVIWVMQDGRIIEQGR
                     YAELSVAGGPFATLLAHRQEEI"
     gene            complement(5083. .6048)
                     /gene="trxB"
                     /locus_tag="S0888"
     CDS             complement(5083. .6048)
                     /gene="trxB"
                     /locus_tag="S0888"
                     /function="enzyme; 2-Deoxyribonucleotide metabolism"
                     /note="residues 1 to 321 of 321 are 95.95 pct identical to
                     residues 1 to 321 of 321 from Escherichia coli K-12 :
                     B0888"
                     /codon_start=1
                     /transl_table=11
                     /product="thioredoxin reductase"
                     /protein_id="AAP16352.1"
                     /db_xref="GI:30040621"
```

```
                     /translation="MGTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQL
                     TTTTEVENWPGDPNDLTGPLLMARMHEHATKFETEIIFDHINKVDLQNRPFRLNGDNG
                     EYTCDALIIATGASARYLGLPSEEAFKGRGVSACATCDGFFYRNQKVAVIGGGNTAVE
                     EALYLSNIASEVHLIHRRDGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGV
                     TGVRLRDTQNSDNIESLDVAGLFVAIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQ
     gene            TSIPGVFAAGDVMDHIYRQAITSAGTGCMAALDAERYLDGLADAK"
                     6592. .7086
                     /gene="lrp"
     CDS             /locus_tag="S0889"
                     6592. .7086
                     /gene="lrp"
                     /locus_tag="S0889"
                     /function="regulator; Central intermediary metabolism:
                     Pool, multipurpose conversions"
                     /note="residues 1 to 164 of 164 are 100.00 pct identical
                     to residues 1 to 164 of 164 from Escherichia coli K-12 :
                     B0889"
                     /codon_start=1
                     /transl_table=11
                     /product="regulator for leucine (or lrp) regulon and
                     high-affinity branched-chain amino acid transport system"
                     /protein_id="AAP16353.1"
                     /db_xref="GI:30040622"
                     /translation="MVDSKKRPGKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTP
                     CLERVRRLERQGFIQGYTALLNPHYLDASLLVFVEITLNRGAPDVFEQFNTAVQKLEE
                     IQECHLVSGDFDYLLKTRVPDMSAYRKLLGETLLRLPGVNDTRTYVVMEEVKQSNRLV
                     IKTR"
     gene            7221. .11249
                     /gene="ftsK"
     CDS             /locus_tag="S0890"
                     7221. .11249
                     /gene="ftsK"
                     /locus_tag="S0890"
                     /function="phenotype; Cell division"
                     /note="residues 1 to 737 of 1342 are 68.24 pct identical
                     to residues 1 to 737 of 1329 from Escherichia coli K-12 :
                     B0890"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein"
                     /protein_id="AAP16354.1"
                     /db_xref="GI:30040623"
                     /translation="MSQEYTEDKEVTLTKLSSGRRLLEALLILIVLFAVWLMAALLSF
                     NPSDPSWSQTAWHEPIHNLGGMPGAWLADTLFFIFGVMAYTIPVIIVGGCWFAWRHQS
                     SDEYIDYFAVSLRIIGVLALILTSCGLAAINADDIWYFASGGVIGSLLSTTLQPLLHS
                     SGGTIALLCVWAAGLTLFTGWSWVTIAEKLGGWILNILTFASNRTRRDDTWVDEDEYE
                     DDEEYEDENHGKQHESRRARILRGALARRKRLAEKFINPMGRQTDAALFSGKRMDDDE
                     EITYTARGVAADPDDVLFSGNRATQPEYDEYDPLLNSAPITEPVAVAAAATTATQSWA
                     APVEPVTQTPPVASVDVPPSQPTVAWQPVPGPQTGEPVIAPAPEGYPQQSQYAQPAVQ
                     YNEPLQQPVQPQQPYYAPAAEQPAQQPYYAPAPEQPVAGNAWQAEEQQSTFAPLSTYQ
                     TEQTYQQPAAQEPLYQQPQPVEQQPVVEPEPVVEETKPARPPLYYFEEVEEKRARERE
                     QLAAWYQPIPEPVKEPEPIKSSLKAPSVAAVPPVEAAAAVSPLASGVKKATLATGAAA
                     TVAAPVFSLANSGGPRPQVKEGIGPQLPRPKRIRVPTRRELASYGIKLPSQRAAEEKA
                     REAQRNQYDSGDQYNDDEIDAMQQDELARQFAQTQQQRYGEQYQHDVPVNAEDADAAA
                     EAELARQFAQTQQQRYSGEQPAGANPFSLDDFEFSPMKALLDDGPHEPLFTPIVEPVQ
                     QPQQPVAPQQQYQQPQQPVAPQQQYQQPQQQVAPQPQYQQPQQPVAPQQQYQQPQQPV
                     APQPQYQQPQQPVAPQPQYQQPQQPVAPQPQDTLLHPLLMRNGDSRPLHKPTTPLPSL
                     DLLTPPPSEVEPVDTFALEQMARLVEARLADFRIKADVVNYSPGPVITRFELNLAPGV
                     KAARISNLSRDLARSLSTVAVRVVEVIPGKPYVGLELPNKKRQTVYLREVLDNAKFRD
```

```
Query Match          73.0%;  Score 332.8;  DB 1;  Length 292504;
Best Local Similarity  83.1%;  Pred. No. 8.6e-87;
Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||
Db     210815 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 210874

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || ||||||||||||| || |||
Db     210875 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 210934

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| |||||||| |  || ||  ||||||||||     ||
Db     210935 TCTTCATTTAATCAGGCAGCCATAATTGGTCAGGCTGGGACTAATAATAGTGCTCAGTTA 210994

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  ||||||||
Db     210995 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAATCGGGCA 211054

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db     211055 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 211114

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||  |||||| | |||||  ||||||| |||     ||||||||||||||||
Db     211115 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 211174

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||||||||||||||||  |
Db     211175 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAGTTGTAGTGCAGAGA 211234

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| ||||| |||||||| || ||||| |||
Db     211235 CAGTCGCAAATGGCAATTCGCGTGACACAACGTTAA 211270
```

```
RESULT 13
AE016759
LOCUS       AE016759               306358 bp    DNA     linear   BCT 24-DEC-2002
DEFINITION  Escherichia coli CFT073 section 5 of 18 of the complete genome.
ACCESSION   AE016759 AE014075
VERSION     AE016759.1  GI:26107527
KEYWORDS    .
SOURCE      Escherichia coli CFT073
  ORGANISM  Escherichia coli CFT073
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Extensive Mosiac Structure Revealed by the Complete Genome Sequence
            of Uropathogenic Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
   PUBMED   12471157
REFERENCE   2  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .306358
                     /organism="Escherichia coli CFT073"
                     /mol_type="genomic DNA"
                     /strain="CFT073"
                     /db_xref="taxon:199310"
     gene            complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
     CDS             complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
                     /note="Residues 27 to 1231 of 1245 are 80.16 pct identical
                     to residues  1 to 1209 of 1218 from GenPept.129 :
                     >emb|CAD05883.1| (AL627276) putative ABC transporter
                     protein [Salmonella enterica subsp. enterica serovar
                     Typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP binding cassette (ABC) transporter homolog"
                     /protein_id="AAN79727.1"
```

```
                /db_xref="GI:26107528"
                /translation="MRDRSVLAFLRNDPASGSYFTDGGGLMPANHTPTPAQSWIVRLA
                RVCWERKKLSVIVVVASVSTILLAALTPLLTRQAVNDALAGNPARLPWLACGLLLIAF
                FDFIGNYVRRGYAGMLSLWVQHTLRGRVFDSIQKLDGAGQDALRTGQVISRTNSDLQQ
                VHTLLQMCPVPLAVFTYYIAGIAVMLWMSPAMTLIVVCVLVCLAITALRARRRVFAQT
                GMASDQLANLTEHIREVLAQISVVKSCVAEMRETHWLDRQSRQIVRVRIGAVISQAMP
                GATMLALPVLGQIVLLCYGGWSVMHGRIDLGTFVAFASFLAMLTGPTRVLASFLVIAQ
                RTQASVERVFALIDTRSQMEDGTESINSQVVGLELENMSFDYHHGDRHILSDISFSLR
                AGETVAVVGASGSGKSTLLMLLARFYDPCSGKIWLNTSEGRQNLRDIRLEALRRRVGI
                VFEDAPLFAGTVAENIAYGHPQATADDIRRAAAAAGASDFINALPKGFDSLLTERGTN
                LSGGQRQRIALARALITAPDVLILDDTTSAVDAVTEAEINTALGRYADEGHMLLVIAR
                RRSTLQLASRVVVLDKGRMVDTGTPAELEARCPAFRALMTGDSDFLATSHNSHNELWP
                AEPATQDDVTDTGDKGFVARMTRVPENAVQQALAGKGRKVTSLLKPVAWMFVIAALLI
                ALDSAAGVGVLILLQHGIDSGVAAGDMSTIGLCALLALCLVIVGWCSYSLQTVFAARA
                AESVQHSVRLRSFGHMLRLGLPWHEKHADSRLTRMTVDVDSLARFLQNGLAGAATSLV
                TMFAIAATMFWLDPILALTALSAVPVAALATMIYRRLSTPAYAQARLEIGKVNSTLQE
                KVSGLRVVQSHGQQELEGARLRALSERFRATRVRAQKYLAVYFPFLTFCTEASYAAVL
                LVGASQVAAGEMTAGVLAAFFLLLGQFYGPVQQLSGIVDAWQQATASGKHIDELLATE
                GTENLGSSSVLPVTGALHLDEVTFSYPDSHEPALNKLTLTIPEGMVVAVVGRSGAGKS
                TLIKLIAGLYFPTHGNIRIGVQMLDDASLTEYRRQIGLVDQDVALFSSDIAENIRYSR
                PSATNEDVEIASQRAGLYEMVCNLPQGFRTPVNNGGADLSAGQRQLIALARAQLANAH
                ILLLDEATSCLDRTSEERLMSSLTDVVHAGKHSALIVAHRLTTAQRCDLIAVIDKGLL
                AEYGTHEQLLSAGGLYTRLWHDSVSSTALHRQHNMKEETPG"
gene            complement(3882. .5045)
                /gene="iroB"
                /locus_tag="c1254"
CDS             complement(3882. .5045)
                /gene="iroB"
                /locus_tag="c1254"
                /note="Escherichia coli O157:H7 ortholog: z1190"
                /codon_start=1
                /transl_table=11
                /product="Putative glucosyltransferase"
                /protein_id="AAN79728.1"
                /db_xref="GI:26107529"
                /translation="MRRLPDLDRQAERDFLMRILFVGPPLYGLLYPVLSLAQAFRVNG
                HEVLIASGGQFAQKAAEAGLVVFDAAPGLDSEAGYRHHEAQRKKSNIGTQMGNFSFFS
                EEMADHLVEFAGHWRPDLIIYPPLGVIGPLIAAKYDIPVVMQTVGFGHTPWHIRGVTR
                SLTDAYRRHNVGATPRDMAWIDVTPPSMSILENDGEPIIPMQYVPYNGGAVWEPWWER
                RPDRKRLLVSLGTVKPMVDGLDLIAWVMDSASEVDAEIILHISANARSDLRSLPSNVR
                LVDWIPMGVPLNGADGFIHHGGAGNTLTALHAGIPQIVFGQGADRPVNARVVAERGCG
                IIPGDVGLSSNMINAFLNNRSLRKASEEVAAEMAAQPCPGEVAKSLITMVQKG"
gene            5535. .5834
                /locus_tag="c1255"
CDS             5535. .5834
                /locus_tag="c1255"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79729.1"
                /db_xref="GI:26107530"
                /translation="MKPTLLQTVENICLQILFVLFLFTGTLIDVEALVSCSSVAKNVQ
                RTPASSVSSSVYSASVSRTAADIMLLSNNPLFAPGQVSFFGCSLAASGIHVTILE"
repeat_region   5867. .6577
                /insertion_seq="IS1541A-like"
gene            5961. .6470
                /locus_tag="c1256"
CDS             5961. .6470
                /locus_tag="c1256"
                /note="Residues 18 to 169 of 169 are 94.07 pct identical
                to residues  1 to 152 of 152 from SwissProt.40 :
                >sp|Q57334|T200_SALTY Transposase for insertion sequence
                element IS200"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79730.1"
                /db_xref="GI:26107531"
                /translation="MRSGNCKCQTRNQKGVPMGNEKSLAHTRWNCKYHIVFAPKYRRQ
                VFYREKRRAIGCILRKLCEWKSVRILEAECCADHIHMLVEIPPKMSVSGFMGYLKGKS
                SLMPYEQFGDLKFKYRNREFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPY
                PGSPFTGRK"
repeat_region   complement(6578. .6940)
                /note="incomplete"
                /insertion_seq="IS4"
gene            complement(6598. .6960)
                /locus_tag="c1257"
CDS             complement(6598. .6960)
                /locus_tag="c1257"
                /function="IS, phage, Tn"
                /note="Residues 8 to 120 of 120 are 98.23 pct identical to
                residues 316 to 428 of 442 from MG1655 : b4278"
                /codon_start=1
                /transl_table=11
                /product="Putative conserved protein"
                /protein_id="AAN79731.1"
                /db_xref="GI:26107532"
                /translation="MPRQPDQGYREIKQTMQLSRLTLRSKKPELVEQELWGVLLAYNL
                VRYQMIKMAEHLKGYWPNQLSFSESCGMVMRMLMTLQGASPGRIPELMRDLASMGQLV
                KLPTRRETAFPRVVKERP"
gene            7651. .8013
                /locus_tag="c1258"
CDS             7651. .8013
                /locus_tag="c1258"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79732.1"
                /db_xref="GI:26107533"
                /translation="MLGACFYSWGTSFYYGFPWWGAELVLMMWPEACFMLLQSSLYSL
                LDGEQVLFSFWELNKKQYEGFEFYQIFSGNIVASCSSCSGVSSNSSAHYARCTDFMRH
                YSLNNYAACQLGRQACLS"
gene            7835. .8269
                /locus_tag="c1259"
CDS             7835. .8269
                /locus_tag="c1259"
                /note="Residues 1 to 139 of 144 are 50.71 pct identical to
                residues  104 to 242 of 258 from GenPept.129 :
                >gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
                [Shigella flexneri]"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79733.1"
                /db_xref="GI:26107534"
                /translation="MKDLSFIRFFLAILLLLVPPALEFPVIHRHITPGALTLCVITAL
                IITLLASLGGRLVSVKCISEISFIRRHRIECMMAGFMIYFWTFSLIAGWYKPQFKKEY
                QTLHYENARYYVLARYDGRLVLSQSYRSGSRKFVIINGGHRL"
gene            8477. .8553
                /locus_tag="c5534"
tRNA            8477. .8553
                /locus_tag="c5534"
                /product="tRNA-Arg"
                /function="RNA; tRNA"
gene            complement(8832. .9290)
                /locus_tag="c1260"
CDS             complement(8832. .9290)
                /locus_tag="c1260"
                /function="putative enzyme; Transposases Insertion
                Sequence Associated"
                /note="Escherichia coli O157:H7 ortholog: z4324"
                /codon_start=1
                /transl_table=11
                /product="Putative Transposase"
                /protein_id="AAN79734.1"
                /db_xref="GI:26107535"
                /translation="MASAADARRIVSHYERRWLIEEYHKAWKSGGTCVESLRMQTRDN
                LERMVVIKAFIAVRMLGLRQEGISEETQNDSCKKILTPTEWKLLWVKLEGKQLPSQTP
                TLKWACLKLGRWHDSKRTGRPGWVVMWDGWFRLQDMVEGYPVMKSLDQEI"
gene            complement(9302. .9667)
                /locus_tag="c1261"
CDS             complement(9302. .9667)
                /locus_tag="c1261"
                /note="Residues 1 to 35 of 121 are 97.14 pct identical to
                residues  59 to 93 of 316 from GenPept.129 :
                >gb|AAK18437.1|AF348706_126 (AF348706) putative
```

```
                     transposase [Shigella flexneri]"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79735.1"
                     /db_xref="GI:26107536"
                     /translation="MAERLGEIQKRVITVCDREADIWHYLHYKVSHGQRFVVRTAQNS
```

```
  Query Match             73.0%;  Score 332.8;  DB 1;  Length 306358;
  Best Local Similarity   83.1%;  Pred. No. 8.6e-87;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db      44773 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 44832

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| |||||||||||||||| || |||
Db      44833 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTTGCGGTAAATGAATTGAGTAAG 44892

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db      44893 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 44952

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db      44953 CGGCAGGGAGGCTCAAAACTTTTGACGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 45012

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | || ||| |||| ||||||| |||| |||||||| ||  |||||| |||||||
Db      45013 AAGATTGATCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 45072

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||     |||||||||||||||||
Db      45073 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 45132

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db      45133 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 45192

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db      45193 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 45228
```

```
RESULT 14
AF275733
LOCUS       AF275733                1711 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
            CsgA protein (csgA) genes, complete cds.
ACCESSION   AF275733
VERSION     AF275733.1  GI:14039399
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Mutations in the csgD promoter associated with variations in curli
            expression in certain strains of Escherichia coli O157:H7
  JOURNAL   Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
  MEDLINE   21218556
   PUBMED   11319125
REFERENCE   2  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
            Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES             Location/Qualifiers
     source          1. .1711
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="43895 Red Variant"
                     /serotype="O157:H7"
                     /db_xref="taxon:562"
     gene            757. .1212
                     /gene="csgB"
     CDS             757. .1212
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="AAK53211.1"
                     /db_xref="GI:14039400"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1253. .1711
                     /gene="csgA"
     CDS             1253. .1711
                     /gene="csgA"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="AAK53212.1"
                     /db_xref="GI:14039401"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
ORIGIN
```

```
  Query Match             72.6%;  Score 331.2;  DB 1;  Length 1711;
  Best Local Similarity   82.9%;  Pred. No. 2.2e-86;
  Matches  378;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db        757 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 816

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| |||||||||||||||| || |||
Db        817 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 876

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db        877 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 936

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db        937 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 996

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db        997 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 1056

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||     |||||||||||||||||
Db       1057 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 1116

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       1117 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 1176

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db       1177 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 1212
```

```
RESULT 15
AE005315
LOCUS       AE005315                10190 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
            of 155.
ACCESSION   AE005315 AE005174
VERSION     AE005315.1  GI:12514572
```

```
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10190
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            617. .1072
                     /gene="csgB"
                     /note="synonym: Z1675"
     CDS             617. .1072
                     /gene="csgB"
                     /function="structural component; Cell exterior
                     constituents: Surface structures"
                     /note="Residues 1 to 151 of 151 are 100.00 pct identical
                     to residues 1 to 151 of 151 from Escherichia coli K-12
                     Strain MG1655: B1041"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAG55787.1"
                     /db_xref="GI:12514573"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1113. .1571
                     /gene="csgA"
                     /note="synonym: Z1676"
     CDS             1113. .1571
                     /gene="csgA"
                     /function="regulator; Cell exterior constituents: Surface
                     structures"
                     /note="Residues 1 to 152 of 152 are 96.71 pct identical to
                     residues 1 to 151 of 151 from Escherichia coli K-12 Strain
                     MG1655: B1042"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin major subunit, coiled surface structures;
                     cryptic"
                     /protein_id="AAG55788.1"
                     /db_xref="GI:12514574"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
     gene            1630. .1962
                     /gene="csgC"
                     /note="synonym: Z1677"
     CDS             1630. .1962
                     /gene="csgC"
                     /function="putative structure; Not classified"
                     /note="Residues 1 to 110 of 110 are 98.18 pct identical to
                     residues 1 to 110 of 110 from Escherichia coli K-12 Strain
                     MG1655: B1043"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="AAG55789.1"
                     /db_xref="GI:12514575"
                     /translation="MNALLLLAALSSQITFNTTQQGDMYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     gene            2083. .2394
                     /gene="Z1678"
     CDS             2083. .2394
                     /gene="Z1678"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 103 of 103 are 98.05 pct identical to
                     residues 1 to 103 of 103 from Escherichia coli K-12 Strain
                     MG1655: B1044"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55790.1"
                     /db_xref="GI:12514576"
                     /translation="MFRPPLNSLMLGSLFFPFIAIAGSTAQGGVIHFYGQIVEPACDV
                     STQSSPVEMNCPQNGSVPGKTYSSKALMSGNVKNAQIASVKVQYLDKQKKLAVMNIEY
                     N"
     gene            2489. .3022
                     /gene="Z1679"
     CDS             2489. .3022
                     /gene="Z1679"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 177 of 177 are 100.00 pct identical
                     to residues 1 to 177 of 177 from Escherichia coli K-12
                     Strain MG1655: B1045"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyprotein"
                     /protein_id="AAG55791.1"
                     /db_xref="GI:12514577"
                     /translation="MKTRIHVVQGDITKLAVDVIVNAANPSLMGGGGVDGAIHRAAGP
                     ALLDACLKVRQQQGDCPTGHAVITLAGDLPAKAVVHTVGPVWRGGEQNEDQLLQDAYL
                     NSLRLVAANSYTSVAFPAISTGVYGYPRAAAAEIAVKTVSEFITRHALPEQVYFVCYD
                     EENAHLYERLLTQQGDE"
     gene            2964. .4445
                     /gene="ymdC"
                     /note="synonym: Z1680"
     CDS             2964. .4445
                     /gene="ymdC"
                     /function="putative enzyme; Not classified"
                     /note="Residues 1 to 493 of 493 are 99.59 pct identical to
                     residues 1 to 493 of 493 from Escherichia coli K-12 Strain
                     MG1655: B1046"
                     /codon_start=1
                     /transl_table=11
                     /product="putative synthase"
                     /protein_id="AAG55792.1"
                     /db_xref="GI:12514578"
                     /translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPGQCGLFPL
                     EKSLDAFAARYRLAEMSEHTLDVQYYIWQDDMSGRLLFSALLAAAKRGVRVRLLLDDN
                     NTPGLDDILRLLDSHPRIEVRLFNPFSFRLLRPLGYITDFSRLNRRMHNKSFTVDGVV
                     TLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWYCKSVSPLQQVLDV
                     PEGEMADRIELPASWHNDAMTHRYLRKMESSPFINHLVDGTLPLIWAKTRLLSDDPAK
                     GEGKAKRHSLLPQRLFDIMGSPSERIDIISSYFVPTRAGVAQLLRMVRKXVKIAILTN
                     SLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREQSSTLHDRGITGNSGASLHAKT
                     FSIDGKTVFIGSFNFDPRSTLLNTEMGFVIESETLAQLIDKRFIQSQYDAAWQLRLDR
                     WGRINWVDRHAKKEIVLKKEPATSFWKRVMVRLASILPVEWLL"
     gene            complement(4453. .5610)
                     /gene="ymdD"
```

```
     CDS             /note="synonym: Z1681"
                     complement(4453. .5610)
                     /gene="ymdD"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 385 of 385 are 99.22 pct identical to
                     residues 1 to 385 of 385 from Escherichia coli K-12 Strain
                     MG1655: B1047"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55793.1"
                     /db_xref="GI:12514579"
                     /translation="MNPVPAQREYFXDSIRAWLMLLGIPFHISLIYSSHTWHVNSAEP
                     SLWLTLFNDFIHSFRMQVFFVISGYFSYMLFLRYPLKKWWKVRVERVGIPMLTAIPLL
                     TLPQFIMLQYVKGKAESWPGLSLYDKYNTLAWELISHLWFLLVLVVMTTLCVWIFKRI
                     RNNLENSDKTNKKFSMVKLSVIFLCLGIGYAVIRRTIFIVYPPILSNGTFNFIVMQTL
                     FYLPFFILGALAFIFPHLKALFTTPSRGCTLAAALAFVAYLLNQRYGSGDAWMYETES
                     VITMVLGLWMVNVVFSFGHRLLNFQSARVTYFVNASLFIYLVHHPLTLFFGAYITPHI
                     TSNWLGFLCGLIFVVGIAIILYEIHLRIPLLKFLFSGKPVVKRENDKAPAR"
     gene            6003. .7538
                     /gene="mdoG"
                     /note="synonym: Z1683"
     CDS             6003. .7538
                     /gene="mdoG"
                     /function="enzyme; Osmotic adaptation"
                     /note="Residues 1 to 511 of 511 are 99.80 pct identical to
                     residues 1 to 511 of 511 from Escherichia coli K-12 Strain
                     MG1655: B1048"
                     /codon_start=1
                     /transl_table=11
                     /product="periplasmic glucans biosynthesis protein"
                     /protein_id="AAG55794.1"
                     /db_xref="GI:12514580"
                     /translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAQSLAGKGYEAPK
                     SNLPSVFRDMKYADYQQIQFNHDKAYWNNLKTPFKLEFYHQGMYFDTPVKINEVTATA
                     VKRIKYSPDYFTFGDVQHDKDTVKDLGFAGFKVLYPINSKDKNDEIVSMLGASYFRVI
                     GAGQVYGLSARGLAIDTALPSGEEFPRFKEFWIERPKPTDKRLTIYALLDSPRATGAY
                     KFVVMPGRDTVVDVQSKIYLRDKVGKLGVAPLTSMFLFGPNQPSPANNYRPELHDSNG
                     LSIHAGNGEWIWRPLNNPKHLAVSSFSMENPQGFGLLQRGRDFSRFEDLDDRYDLRPS
                     AWVTPKGEWGKGSVELVEIPTNDETNDNIVAYWTPDQLPEPGKEMNFKYTITFSRDED
                     KLHAPDNAWVQQTRRSTGDVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDTPVTAQTS
                     IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDAKKTTEMRAALVNADQTLSETWSYQLP
                     ANE"
     gene            7531. .10074
                     /gene="mdoH"
                     /note="synonym: Z1684"
     CDS             7531. .10074
                     /gene="mdoH"
                     /function="enzyme; Osmotic adaptation"

  Query Match          72.6%;  Score 331.2;  DB 1;  Length 10190;
  Best Local Similarity  82.9%;  Pred. No. 2.3e-86;
  Matches  378;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||
Db      617 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 676

Qy       61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
            ||  ||  ||||||| | |||  |||||||||||| || |||||||||||| || ||| |
Db      677 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 736

Qy      121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
            |||||||||||||||| ||||| |||||||||| || ||  |||||||||| ||
Db      737 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 796

Qy      181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
            || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db      797 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 856

Qy      241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
            ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| ||||||
Db      857 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 916

Qy      301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
            |||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db      917 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 976

Qy      361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
            || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db      977 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 1036

Qy      421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
            |||||||| |||||||||||||| || ||||| |||
Db     1037 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 1072

Search completed: March 18, 2004, 15:02:27
Job time : 2314.32 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64618	Aac64618 Salmonell
	2	332.8	73.0	456	3	AAC64620	Aac64620 Escherich
	3	74.4	16.3	100	7	ACD68810	Acd68810 E. coli K
	4	66.4	14.6	100	7	ACD68812	Acd68812 E. coli K
	5	64.8	14.2	100	7	ACD68811	Acd68811 E. coli K
	6	37.6	8.2	456	3	AAC64623	Aac64623 AgfA::PT3
c	7	37.6	8.2	2000	7	ADA71938	Ada71938 Rice gene
	8	35.8	7.9	39	3	AAC64608	Aac64608 AgfA (SEF
	9	35	7.7	738	6	AAD32400	Aad32400 Chlamydia
	10	35	7.7	105325	6	ABK94407	Abk94407 DNA encod
	11	34.4	7.5	2006	9	ADC10081	Adc10081 Human NOV
c	12	34.4	7.5	18540	8	ACD26496	Acd26496 Human tum
	13	34.2	7.5	456	3	AAC64629	Aac64629 AgfA::PT3
	14	34	7.5	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	34	7.5	361	2	AAT74141	Aat74141 Salmonell
	16	34	7.5	456	2	AAQ87467	Aaq87467 AgfA sequ
	17	34	7.5	456	2	AAT74142	Aat74142 Salmonell
	18	34	7.5	456	3	AAC64628	Aac64628 AgfA::PT3
	19	34	7.5	456	3	AAC64622	Aac64622 AgfA::PT3
	20	34	7.5	456	3	AAC64617	Aac64617 Salmonell
	21	34	7.5	456	3	AAC64625	Aac64625 AgfA::PT3
	22	34	7.5	456	3	AAC64624	Aac64624 AgfA::PT3
	23	34	7.5	2000	7	ADA71938	Ada71938 Rice gene
c	24	33.8	7.4	203654	7	ABX16034	Abx16034 Human gen
c	25	33.4	7.3	53585	2	AAX20251	Aax20251 Borrelia
	26	33.2	7.3	535	9	ADD29528	Add29528 Mouse cae
	27	33	7.2	33	3	AAC64614	Aac64614 S. enteri
	28	33	7.2	3699	6	ABN70306	Abn70306 Streptoco
	29	33	7.2	3699	6	ABN70834	Abn70834 Streptoco
	30	33	7.2	3717	6	ABN66453	Abn66453 Streptoco
	31	33	7.2	110000	6	ABN71527_03	Continuation (4 of
	32	32.8	7.2	418	3	AAH30547	Aah30547 Human col
	33	32.8	7.2	1083	9	ADC10077	Adc10077 Human NOV
	34	32.8	7.2	1098	9	ADC10079	Adc10079 Human NOV
	35	32.8	7.2	1105	9	ADC10075	Adc10075 Human NOV
	36	32.8	7.2	1168	9	ADC10069	Adc10069 Human NOV
	37	32.8	7.2	1171	9	ADC10071	Adc10071 Human NOV
	38	32.8	7.2	1299	9	ADC10073	Adc10073 Human NOV
	39	32.8	7.2	1938	9	ADC10067	Adc10067 Human NOV
	40	32.8	7.2	2649	7	ADA53932	Ada53932 Human cod
	41	32.8	7.2	2843	7	ACC44405	Acc44405 Gene enco
	42	32.4	7.1	779	1	AAN40272	Aan40272 Sequence
	43	32.4	7.1	1914	4	AAS51585	Aas51585 Staphyloc
	44	32.4	7.1	1929	4	AAS54408	Aas54408 Staphyloc
	45	32.4	7.1	10732	3	AAA10594	Aaa10594 Gene enco

ALIGNMENTS

RESULT 1
AAC64618
ID AAC64618 standard; DNA; 456 BP.
XX
AC AAC64618;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis AgfB DNA sequence SEQ ID NO:2.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36342.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 149 A; 87 C; 109 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.6e-135;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
```

RESULT 2
AAC64620
ID AAC64620 standard; DNA; 456 BP.
XX
AC AAC64620;
XX
DT 26-FEB-2001 (first entry)
XX
DE Escherichia coli CsgB DNA sequence SEQ ID NO:4.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Escherichia coli.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36344.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 146 A; 78 C; 113 G; 119 T; 0 U; 0 Other;

Query Match 73.0%; Score 332.8; DB 3; Length 456;
Best Local Similarity 83.1%; Pred. No. 1.5e-95;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||| ||||| ||||||||||  || ||  ||||||||||  ||
Db        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
```

```
                ||||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
                || ||||| || |||||||| ||||| ||||| || ||||| ||| ||||||||||| |
Db        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
                |||||||| |||||||||||||| || ||||| |||
Db        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456


RESULT 3
ACD68810
ID   ACD68810 standard; DNA; 100 BP.
XX
AC   ACD68810;
XX
DT   18-SEP-2003  (first entry)
XX
DE   E. coli K12 MG1655 biochip probe SEQ ID 80.
XX
KW   Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS   Escherichia coli.
XX
PN   EP1260592-A1.
XX
PD   27-NOV-2002.
XX
PF   17-MAY-2001; 2001EP-00112179.
XX
PR   17-MAY-2001; 2001EP-00112179.
XX
PA   (MWGB-) MWG-BIOTECH AG.
XX
PI   Donner H,  Drescher B,  Huber A,  Weber J;
XX
DR   WPI; 2003-241155/24.
XX
PT   Biochip containing probes complementary with open reading frames in
PT   Escherichia coli K12, useful for detecting gene expression and expression
PT   patterns.
XX
PS   Claim 3; Page 23; 2004pp; German.
XX
CC   This invention describes a novel biochip comprising probe spots, each
CC   containing many identical probes. The probes are nucleotide sequences of
CC   30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC   least one includes a segment of at least 20 bases identical with, or
CC   complementary to, a segment of an open reading frame (orf) of Escherichia
CC   coli K12. The biochip is used for specific detection of gene expression
CC   in K12 and for determining the gene expression pattern, e.g. for
CC   diagnostic determination of which E. coli strains are present in the gut,
CC   and to determine the effects of e.g. growth media on gene expression. The
CC   biochip provides as comprehensive as possible detection of the K12
CC   genome, with simultaneous analysis of many different genes with a single
CC   device, and comparison of gene expression between K12 and its mutants or
CC   other E. coli strains in a single experiment. Apart from qualitative and
CC   quantitative information about gene expression, it also allows
CC   measurements of population densities for the various strains. The use of
CC   synthetic oligonucleotides for preparation of probes allows free
CC   variation in probe length and ensures high purity (and thus selectivity,
CC   reactivity and reproducibility); also synthetic probes are generally
CC   shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC   ACD81540 represent oligonucleotide probes used with the biochip described
CC   in the invention
XX
SQ   Sequence 100 BP; 27 A; 19 C; 26 G; 28 T; 0 U; 0 Other;

  Query Match             16.3%;  Score 74.4;  DB 7;  Length 100;
  Best Local Similarity   84.0%;  Pred. No. 1.6e-13;
  Matches   84;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        266 ACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTT 325
              || |||| |||||||| ||  |||||| |||||||||||||||||| |||||| | ||||
Db          1 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 60

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAA 365
              | ||||||| |||    |||||||||||||||||||| ||
Db         61 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 100


RESULT 4
ACD68812
ID   ACD68812 standard; DNA; 100 BP.
XX
AC   ACD68812;
XX
DT   18-SEP-2003  (first entry)
XX
DE   E. coli K12 MG1655 biochip probe SEQ ID 82.
XX
KW   Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS   Escherichia coli.
XX
PN   EP1260592-A1.
XX
PD   27-NOV-2002.
XX
PF   17-MAY-2001; 2001EP-00112179.
XX
PR   17-MAY-2001; 2001EP-00112179.
XX
PA   (MWGB-) MWG-BIOTECH AG.
XX
PI   Donner H,  Drescher B,  Huber A,  Weber J;
XX
DR   WPI; 2003-241155/24.
XX
PT   Biochip containing probes complementary with open reading frames in
PT   Escherichia coli K12, useful for detecting gene expression and expression
PT   patterns.
XX
PS   Claim 3; Page 23; 2004pp; German.
XX
CC   This invention describes a novel biochip comprising probe spots, each
CC   containing many identical probes. The probes are nucleotide sequences of
CC   30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC   least one includes a segment of at least 20 bases identical with, or
CC   complementary to, a segment of an open reading frame (orf) of Escherichia
CC   coli K12. The biochip is used for specific detection of gene expression
CC   in K12 and for determining the gene expression pattern, e.g. for
CC   diagnostic determination of which E. coli strains are present in the gut,
CC   and to determine the effects of e.g. growth media on gene expression. The
CC   biochip provides as comprehensive as possible detection of the K12
CC   genome, with simultaneous analysis of many different genes with a single
CC   device, and comparison of gene expression between K12 and its mutants or
CC   other E. coli strains in a single experiment. Apart from qualitative and
CC   quantitative information about gene expression, it also allows
CC   measurements of population densities for the various strains. The use of
CC   synthetic oligonucleotides for preparation of probes allows free
CC   variation in probe length and ensures high purity (and thus selectivity,
CC   reactivity and reproducibility); also synthetic probes are generally
CC   shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC   ACD81540 represent oligonucleotide probes used with the biochip described
CC   in the invention
XX
SQ   Sequence 100 BP; 28 A; 18 C; 27 G; 27 T; 0 U; 0 Other;

  Query Match             14.6%;  Score 66.4;  DB 7;  Length 100;
  Best Local Similarity   79.0%;  Pred. No. 5.6e-11;
  Matches   79;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        123 TTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACG 182
```

```
Db          1 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 60

Qy        183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAA 222
              |||| ||| |||||||| ||| | ||| || | ||||||
Db         61 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 100
```

RESULT 5
ACD68811
ID ACD68811 standard; DNA; 100 BP.
XX
AC ACD68811;
XX
DT 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 81.
XX
KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EP1260592-A1.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MWGB-) MWG-BIOTECH AG.
XX
PI Donner H, Drescher B, Huber A, Weber J;
XX
DR WPI; 2003-241155/24.
XX
PT Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX
PS Claim 3; Page 23; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SQ Sequence 100 BP; 31 A; 20 C; 29 G; 20 T; 0 U; 0 Other;

```
  Query Match           14.2%;  Score 64.8;  DB 7;  Length 100;
  Best Local Similarity  78.0%;  Pred. No. 1.8e-10;
  Matches   78;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        208 GTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAAC 267
              ||| || | |||||||| | | || ||||| || | |||||| |||| ||||||||
Db          1 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 60
```

```
Qy        268 TTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCA 307
              |||| |||||||| ||  |||||| ||||||||||||||
Db         61 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 100
```

RESULT 6
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

```
  Query Match             8.2%;  Score 37.6;  DB 3;  Length 456;
  Best Local Similarity   51.2%;  Pred. No. 0.15;
  Matches   88;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
              ||   |||   || |  ||  |||  ||||  |||  | ||   ||  | ||  |   ||
Db        203 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 262

Qy        332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTA 391
               || |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACG 322

Qy        392 CGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGT 443
              |  | ||  | |   ||   ||      || |    ||  ||| ||  || ||
Db        323 CTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGT 374


RESULT 7
ADA71938/c
ID   ADA71938 standard; DNA; 2000 BP.
XX
AC   ADA71938;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Rice gene, SEQ ID 5263.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
OS   Oryza sativa.
XX
PN   WO2003000898-A1.
XX
PD   03-JAN-2003.
XX
PF   22-JUN-2001; 2001WO-IB001105.
XX
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H,  Chen W,  Cooper B,  Glazebrook J,  Goff SA,  Hou Y;
PI   Katagiri F,  Quan S,  Tao Y,  Whitham S,  Xie Z,  Zhu T,  Zou G;
XX
DR   WPI; 2003-175290/17.
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction. (M1) is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

  Query Match             8.2%;  Score 37.6;  DB 7;  Length 2000;
  Best Local Similarity   12.8%;  Pred. No. 0.26;
  Matches   57;  Conservative  195;  Mismatches  189;  Indels    3;  Gaps    2;

Qy         14 TGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACAAATTATG 73
              |||    |  |::::|: ::: : :   ::  :::|  : |::  : :: :|:|:   ::

Db       1027 TGTAGGTGCTKMWRTYTRSMSWTYAMWKKYTKYMTAYSSTWKWYWAYKWRAYAWSRSRKT 968

Qy         74 ATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATC 133
              ::|||| : |:|   ::|  :: :      :      :::: :|:: :::  :  :
Db        967 WWCTGGKRMATYCGTKMWAAGRWRWRMAWCWYCCMWKWWKMTSCMWWKYWRTWSCWYTMW 908

Qy        134 AGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGAT 193
              :| ::  :   :  : |: ::: : ::    :|:|: | :|:   ::    | :: : |
Db        907 WGAMRYAYYAMRRRRWTYKWSWRRMYWTMTKWAWTWMTCMCMAKWYMATGWATWMWWRYT 848

Qy        194 CAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGG 253
              :    : :|  :|  : | |: :  :  | :| :::: |:   :  :  ::   :  : :
Db        847 MYTYCYAMTCAKCKYKMAMTKWWTTWACAWRATSWRWRAMAGMRWKRYKMKRAYWWRWWR 788

Qy        254 CAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATAT 313
              |:: |:| ::: :: ::  ::||:::    |:    : ::::::::: :  | ::  |:::
Db        787 CWKAGWARWMKSRYRWKWKKYATRYYWKMWAMTWWWSWRRWKSYRMWSGMGRMRWSAWRY 728

Qy        314 CGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCA 373
              |:  :::|  |:  :: ::| : :    ::  ::  :: :  :: :::: ::|:|: : :
Db        727 CS--RMKCAKTKYASSARWTKRAKRSYRYRRRWYWKRKGWTYRYRYWRSCRMTRARMSKR 670

Qy        374 ATATTACCCAGTACGGTACGCA-GAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATG 432
              :::  |::   ::: : | ::: :    :|:| ::|:::  ::::  ::  :::: :
Db        669 RKWAGASMKSCWMYWRGARSMWYSKYSCSAKCCKKTRYMTSSYMSTGMYGMYSSYKSMSW 610

Qy        433 GCTATTCGCGTCACCCAACGCTAA 456
               ::: ::|  || ::  : |:|::
Db        609 TSKMSYMGKMTCTMYTSMKGSTRR 586


RESULT 8
AAC64608
ID   AAC64608 standard; DNA; 39 BP.
XX
AC   AAC64608;
XX
DT   06-AUG-2003  (revised)
DT   26-FEB-2001  (first entry)
XX
DE   AgfA (SEF17) recombinant agfA::PT3a generating PCR primer 17-A.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; PCR primer; ss.
XX
OS   Salmonella enteritidis.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Example 2; Page 61; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
```

```
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC   correct OS field.)
XX
SQ   Sequence 39 BP; 13 A; 7 C; 10 G; 9 T; 0 U; 0 Other;

  Query Match             7.9%;  Score 35.8;  DB 3;  Length 39;
  Best Local Similarity   94.9%;  Pred. No. 0.21;
  Matches   37;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        393 GCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATAT 431
              ||||||  |||||||||||||||||||||||||||||||
Db          1 GCAGAATTCAGCAGTTGTAGTGCAGAAACAGTCGCATAT 39


RESULT 9
AAD32400
ID   AAD32400 standard; DNA; 738 BP.
XX
AC   AAD32400;
XX
DT   29-AUG-2003  (revised)
DT   18-JUN-2002  (first entry)
XX
DE   Chlamydia pneumoniae nlpD gene.
XX
KW   Chlamydiaceae family; chronic infection; persistent infection; pyk; nlpD;
KW   Cpn0585; regulatory pathway; biosynthetic pathway; ompA; ompB; hsp60;
KW   lipopolysaccharide; cardiovascular system; respiratory tract; therapy;
KW   genital tract; reproductive system; atherosclerotic tissue; macrophage;
KW   multiple sclerosis; conjunctiva; prophylaxis; antibacterial; gene; ds.
XX
OS   Chlamydophila pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .738
FT                   /*tag=  a
FT                   /product= "Chlamydia pneumoniae nlpD protein"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200214516-A1.
XX
PD   21-FEB-2002.
XX
PF   17-AUG-2001; 2001WO-AU001021.
XX
PR   18-AUG-2000; 2000AU-00009540.
XX
PA   (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
PA   (MATH/) MATHEWS S A.
XX
PI   Timms P;
XX
DR   WPI; 2002-269197/31.
DR   P-PSDB; AAE20289.
XX
PT   Detecting Chlamydial organism in its persistent phase by detecting
PT   expression change of range of genes belonging to their respective
PT   biosynthetic pathways when expression is compared to that of organism in
PT   lytic phase.
XX
PS   Disclosure; Page 117-118; 196pp; English.
XX
CC   The invention relates to composition and methods for detecting organisms
CC   of the Chlamydiaceae family, including species of Chlamydophila and
CC   Chlamydia, in the persistent phase of their developmental cycle and for
CC   the diagnosis of chronic or persistent infections caused by such
CC   organisms. The composition is useful for modulating the expression of
CC   gene such as pyk, nlpD, Cpn0585, a gene belonging to same regulatory/
CC   biosynthetic pathway and ompA, ompB, hsp60, a gene involved in
CC   lipopolysaccharide biosynthesis. It is also useful for modulating the the
CC   level and/or functional activity of an expression product of these genes,
CC   where the gene is present in an epithelial cell (selected from
CC   cardiovascular system, respiratory tract, genital tract, reproductive
CC   system or conjunctiva), macrophage, or a cell associated with
CC   atherosclerotic tissue or associated with multiple sclerosis brain
CC   tissue. The composition is useful for treatment and/or prophylaxis of a
CC   chronic infection caused by an organism of the Chlamydiaceae family in a
CC   patient. Antigen associated with the persistent phase of the
CC   developmental cycle of an organism of the Chlamydiaceae family, is useful
CC   in the manufacture of a medicament, for treating and/or preventing
CC   Chlamydiaceae infection in a patient. The present sequence is Chlamydia
CC   pneumoniae nlpD gene. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ   Sequence 738 BP; 231 A; 158 C; 166 G; 183 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 35;  DB 6;  Length 738;
  Best Local Similarity   57.9%;  Pred. No. 1.2;
  Matches   62;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        340 GCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAA 399
              |||  ||   ||||  |  |  ||  |  |  |||  |||||  |  |      |  |  |  ||
Db          1 GCTGTTGATCAGAGAAATGCAGGCAGTGAGGTAAATATGAATCGTAGAGACATGGTAATA 60

Qy        400 ACAGCAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCAC 446
              |||||  ||  ||||||  |        |  ||  |  |||||  ||   ||||||
Db         61 ACAGCTGTCGTAGTGAATGCTATATTGCTTGTGGCTCTTTTCGTCAC 107


RESULT 10
ABK94407
ID   ABK94407 standard; DNA; 105325 BP.
XX
AC   ABK94407;
XX
DT   27-AUG-2002  (first entry)
XX
DE   DNA encoding endothelin-3 (EDN-3).
XX
KW   Endothelin 3; EDN-3; endothelin converting enzyme; ECE;
KW   endothelin receptor; EDNR; signaling system; cardiovascular disease;
KW   coronary heart disease; hypertension; atherosclerosis; angiogenesis;
KW   fatty acid metabolism; diabetes; familial hypercholesterolaemia;
KW   forensic marker; transgenic animal; solid support; SNP;
KW   cardiovascular regulator; gene; ds; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(40215,A)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(59430,C)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63843,T)
```

```
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   WO200224747-A2.
XX
PD   28-MAR-2002.
XX
PF   31-AUG-2001; 2001WO-EP010087.
XX
PR   19-SEP-2000; 2000EP-00120123.
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI   Brinkmann U,  Hoffmeyer S;
XX
DR   WPI; 2002-435060/46.
XX
PT   Novel polynucleotide of the endothelin/endothelin converting
PT   enzyme/receptors of endothelin and endothelin converting enzyme signaling
PT   system associated with cardiovascular disease, useful for treating the
PT   disease.
XX
PS   Claim 1; Page; 190pp; English.
XX
CC   The invention describes a polynucleotide (I) of the endothelin
CC   (EDN)/endothelin converting enzyme (ECE)/receptors of EDN and ECE (EDNR)
CC   signaling system which is associated with a cardiovascular disease. (I),
CC   the gene encoding EDN, ECE or EDNR (II) or a vector (III) expressing (I)
CC   or (II) is useful for producing cells capable of expressing a molecular
CC   variant polypeptide which is associated with a cardiovascular disease.
CC   (II), (III), the EDN, ECE or EDNR polypeptide, or a cell expressing a
CC   molecular variant gene comprising (I) is useful for identifying and
CC   obtaining a pro-drug or drug capable of modulating the activity of a
CC   molecular variant of a polypeptide of the EDN/EDNR/ECE signaling system
CC   or its gene product, or for identifying and obtaining an inhibitor of the
CC   activity of a molecular variant of a polypeptide of the EDN/EDNR/ECE
CC   signaling system or its gene product. The isolated proteins and
CC   polynucleotides encoding them are useful for preparation of a
CC   pharmaceutical composition for treating a cardiovascular disease such as
CC   coronary heart disease, hypertension, atherosclerosis, or related to
CC   abnormal angiogenesis or fatty acid metabolism e.g. diabetes and familial
CC   hypercholesterolaemia. The gene or a polynucleotide fragment of the
CC   EDN/ECE/EDNR signaling system are useful as forensic markers, for
CC   creating a transgenic animal and in creation of a solid support
CC   comprising polynucleotides, genes, vectors, polypeptides, antibodies or
CC   host cells of the invention. This sequence encodes the cardioavscular
CC   regulator Endothelin-3 (EDN-3). Note: This sequence does not appear in
CC   the specification but has been obtained from GenBank using information
CC   given in the invention
XX
SQ   Sequence 105325 BP; 26263 A; 24618 C; 25835 G; 28609 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 35;  DB 6;  Length 105325;
  Best Local Similarity   56.5%;  Pred. No. 8.2;
  Matches   65;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||  || |       || ||  ||||   |  || || |||  |  ||  |||
Db      66519 ATGAAAAAGCAAATTAAGAGCATAATCCTGACTGCATGGGATGTGCCCAGCCTTCTAACA 66578

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAA 115
              |  ||||| | |   |||||   |||| | |||   ||        ||| ||||||
Db      66579 GGAACAAACTCTTTCCTGGCCTTTTCAAAGTATCTGTTACTATGTAATAAATTAA 66633


RESULT 11
ADC10081
ID   ADC10081 standard; DNA; 2006 BP.
XX
AC   ADC10081;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human NOVX polypeptide coding sequence SEQ ID NO: 101.
XX
KW   ds; gene; cytostatic; antidiabetic; anorectic; cerebroprotective;
KW   neuroprotective; antiinflammatory; gene therapy; antisense therapy;
KW   thyromimetic; NOVX; pathology; cancer; diabetes; obesity;
KW   endocrine disorder; CNS disorder; inflammatory disorder;
KW   chromosome mapping; tissue typing; predictive medicine.
XX
OS   Homo sapiens.
XX
PN   WO2003000842-A2.
XX
PD   03-JAN-2003.
XX
PF   04-JUN-2002; 2002WO-US017443.
XX
PR   04-JUN-2001; 2001US-0295607P.
PR   04-JUN-2001; 2001US-0295661P.
PR   06-JUN-2001; 2001US-0296404P.
PR   06-JUN-2001; 2001US-0296418P.
PR   07-JUN-2001; 2001US-0296575P.
PR   11-JUN-2001; 2001US-0297414P.
PR   12-JUN-2001; 2001US-0295573P.
PR   12-JUN-2001; 2001US-0297567P.
PR   14-JUN-2001; 2001US-0298285P.
PR   15-JUN-2001; 2001US-0298528P.
PR   18-JUN-2001; 2001US-0299133P.
PR   19-JUN-2001; 2001US-0299230P.
PR   21-JUN-2001; 2001US-0299949P.
PR   22-JUN-2001; 2001US-0300177P.
PR   26-JUN-2001; 2001US-0300883P.
PR   28-JUN-2001; 2001US-0301530P.
PR   28-JUN-2001; 2001US-0301550P.
PR   03-JUL-2001; 2001US-0302951P.
PR   31-JUL-2001; 2001US-0308890P.
PR   14-SEP-2001; 2001US-0322297P.
PR   25-SEP-2001; 2001US-0324669P.
PR   03-DEC-2001; 2001US-0337477P.
PR   14-DEC-2001; 2001US-0341562P.
PR   21-FEB-2002; 2002US-0358656P.
PR   21-FEB-2002; 2002US-0359122P.
PR   22-FEB-2002; 2002US-0358978P.
PR   22-FEB-2002; 2002US-0359034P.
PR   22-FEB-2002; 2002US-0359035P.
PR   22-FEB-2002; 2002US-0359121P.
PR   27-FEB-2002; 2002US-0359964P.
PR   01-MAR-2002; 2002US-0360858P.
PR   12-MAR-2002; 2002US-0363430P.
PR   12-MAR-2002; 2002US-0363676P.
PR   10-APR-2002; 2002US-0371346P.
PR   10-MAY-2002; 2002US-0379444P.
PR   04-JUN-2002; 2002US-00379444.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML,  Anderson DW,  Berghs C,  Casman SJ,  Catterton E;
PI   Dipippo VA,  Edinger SR,  Eisen A,  Ellerman K,  Gangolli EA;
PI   Gerlach VL,  Gorman L,  Guo X,  Herrmann JL,  Hjalt T,  Ji W,  Kekuda R;
PI   Khramtsov NV,  Li L,  Liu X,  Malyankar UM,  Miller CE,  Millet I;
PI   Ort T,  Padigaru M,  Patturajan M,  Pena CEA,  Rastelli L,  Rieger DK;
PI   Rothenberg ME,  Shenoy SG,  Shimkets RA,  Smithson G,  Spaderna SK;
PI   Spytek KA,  Stone DJ,  Vernet CAM,  Zhong H,  Zhong M,  Alsobrook JP;
PI   Burgess CE,  Lepley DM;
XX
DR   WPI; 2003-210149/20.
DR   P-PSDB; ADC10082.
XX
PT   New isolated NOVX polypeptides and nucleic acid molecules useful for
PT   treating, preventing and diagnosing pathological conditions with NOVX-
PT   associated disorders, such as cancer, obesity, diabetes and inflammatory
PT   or CNS diseases.
XX
PS   Claim 20; SEQ ID NO 101; 772pp; English.
```

```
XX
CC   The invention relates to novel isolated polypeptides, mature form of the
CC   polypeptide, a sequence that is 95% identical to the polypeptide or the
CC   polypeptide comprising one or more conservative substitutions. The NOVX
CC   polypeptide is useful for treating or preventing a pathology associated
CC   with the polypeptide e.g. disorders associated with aberrant expression
CC   or activity of the polypeptide, such as cancer, diabetes, obesity, and
CC   endocrine, CNS and inflammatory disorders. They can also be used in
CC   various detection and screening assays, chromosome mapping, tissue typing
CC   and predictive medicine. This sequence corresponds to the coding sequence
CC   for one of the polypeptides of the invention.
XX
SQ   Sequence 2006 BP; 637 A; 390 C; 493 G; 486 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34.4;  DB 9;  Length 2006;
  Best Local Similarity   48.9%;  Pred. No. 2.7;
  Matches   92;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        208 GTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAAC 267
              ||| ||   | || |||||  |   |    |  | ||   ||  | ||| |  | |  ||
Db        540 GTTCTTGATCTAGCAGGAGTGATACAAAACGTCATAGATCACGTGCCAGAGCCTCACTAC 599

Qy        268 TTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTAC 327
              | |    | | ||  |  | |   || ||   || | ||    |   |   | |  | |
Db        600 CTGGTTGCTGTGGATGAGTCTGTTCATACCCTGGAAGACATAGTCAAGTGTATCAGTAAA 659

Qy        328 GGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTAC 387
              || | |  | | |  |||||||||    | || | || ||  |  | ||| || ||
Db        660 AATACTGGCCCTGGGAAAATCCAGAAAATACCCAGAGAAAATGCATACCTAACCAAGGAC 719

Qy        388 GGTACGCA 395
                  |||||
Db        720 TTAACGCA 727


RESULT 12
ACD26496/c
ID   ACD26496 standard; DNA; 18540 BP.
XX
AC   ACD26496;
XX
DT   09-SEP-2003  (first entry)
XX
DE   Human tumour necrosis factor receptor related protein-1, TRP-1, DNA.
XX
KW   Human; ds; gene; TRP-1; inflammatory disorder; rheumatoid arthritis;
KW   tumour necrosis factor receptor related protein-1; Crohn's disease; AIDS;
KW   asthma; irritable bowel syndrome; ulcerative colitis; multiple sclerosis;
KW   diabetes mellitus; immune disorder; autoimmune disease; allergy; stroke;
KW   systemic lupus erythematosus; neurodegenerative disorder; prostatitis;
KW   Alzheimer's disease; Parkinson's disease; cell proliferative disorder;
KW   cell differentiation disorder; ischaemic injury; myocardial infarction;
KW   apoptotic disorder; gastrointestinal and reproductive tract disorder;
KW   cancer; ulcer; bone disorder; osteoporosis; blood disorder; hepatitis;
KW   aplastic anaemia; thrombocytopenia; viral disorder; viral encephalitis;
KW   chromosome Xq12.
XX
OS   Homo sapiens.
XX
PN   US2003059889-A1.
XX
PD   27-MAR-2003.
XX
PF   28-AUG-2002; 2002US-00231426.
XX
PR   05-SEP-2001; 2001US-0317151P.
XX
PA   (BODI/) BODIAN D L.
PA   (LABO/) LABOW M A.
PA   (MICK/) MICKANIN C S.
XX
PI   Bodian DL,  Labow MA,  Mickanin CS;
XX
DR   WPI; 2003-521912/49.
DR   P-PSDB; ABU62739.
XX
PT   New tumor necrosis factor receptor related protein-1 polypeptide and
PT   nucleic acid, useful for diagnosing and treating disorders in a human,
PT   e.g. an inflammatory disorder, an immune disorder or a neurodegenerative
PT   disorder.
XX
PS   Claim 16; Page 21-30; 36pp; English.
XX
CC   The invention relates to a new isolated tumour necrosis factor receptor
CC   related protein-1 polypeptide, TRP-1. The tumour necrosis factor receptor
CC   related protein-1 polypeptide, nucleic acid and antibody are useful for
CC   diagnosing and treating disorders in a human, such as an inflammatory
CC   disorder (e.g. rheumatoid arthritis, Crohn's disease, AIDS, allergies,
CC   asthma, irritable bowel syndrome, ulcerative colitis, multiple sclerosis
CC   or diabetes mellitus), an immune disorder (e.g. autoimmune disease or
CC   systemic lupus erythematosus), a neurodegenerative disorder (e.g.
CC   Alzheimer's disease or Parkinson's disease), a cell proliferative
CC   disorder (e.g. various types of cancer), a cell differentiation disorder
CC   (e.g. stroke, ischaemic injury or myocardial infarction), an apoptotic
CC   disorder, a gastrointestinal and reproductive tract disorder (e.g. ulcers
CC   or prostatitis), bone disorder (e.g. osteoporosis), blood disorder (e.g.
CC   aplastic anaemia or thrombocytopenia) or viral disorder (e.g. hepatitis
CC   or viral encephalitis). The polypeptides are also useful in manufacturing
CC   a medicament for treating the disorders. The methods are useful for
CC   identifying molecules that bind to and/or modulate the activity of the
CC   polypeptide, or the transcription or translation of the nucleic acid. The
CC   present sequence represents human tumour necrosis factor receptor related
CC   protein-1, TRP-1, DNA which is located on chromosome Xq12
XX
SQ   Sequence 18540 BP; 5173 A; 3532 C; 3591 G; 6244 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34.4;  DB 8;  Length 18540;
  Best Local Similarity   52.0%;  Pred. No. 6.5;
  Matches   77;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         86 CAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTA 145
              ||||   ||||| |     | |  |    || || | |  |   |  | ||| |   |||
Db       5513 CAGATGGTAATTATTAAAAATAAAATGACAGAAATTTTAAAAATTCACTGGATGGGCTTA 5454

Qy        146 TTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGT 205
              | ||  ||  |  || ||| ||||   ||| | ||   ||||  || ||| |  |||
Db       5453 GTAGTAGAGAAGAGACAGATGATAGAATCAGTGAACTTGAGGACAGAACAACAGAATTTA 5394

Qy        206 CCGTTATTTCACAAGAAGGAGGAAATAA 233
                  |  |||| | |||| |||| |
Db       5393 TGCAATATGAACAACAGGGAGAAAATGA 5366


RESULT 13
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
```

```
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34.2;  DB 3;  Length 456;
  Best Local Similarity   56.8%;  Pred. No. 1.8;
  Matches   63;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        275 ATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATA 334
              | || |  ||  | ||  ||| |||| |||  | ||   ||  | ||  |   || ||
Db        206 AAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTA 265

Qy        335 GTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
                |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        266 CTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34;  DB 2;  Length 361;
  Best Local Similarity   56.1%;  Pred. No. 1.9;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
              ||   |||   || |  ||  |||  ||||  |||  | ||   ||  | ||  |   ||
Db        140 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 199

Qy        332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
                || |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        200 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 253


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
```

```
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag= b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag= c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag= d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag= e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34;  DB 2;  Length 361;
  Best Local Similarity   56.1%;  Pred. No. 1.9;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
              ||   |||   || |  ||  |||  ||||  |||  | ||   ||  | ||  |   ||
Db        140 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 199

Qy        332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              || |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        200 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 253


Search completed: March 18, 2004, 09:11:47
Job time : 290.405 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	41	9.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	34	7.5	361	1	US-08-233-788A-56	Sequence 56, Appl
	3	34	7.5	456	1	US-08-233-788A-58	Sequence 58, Appl
	4	32.2	7.1	1443	4	US-09-328-352-2823	Sequence 2823, Ap
c	5	32.2	7.1	1830121	4	US-09-557-884-1	Sequence 1, Appli
c	6	32.2	7.1	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c	7	32	7.0	651	4	US-09-134-000C-1323	Sequence 1323, Ap
	8	31.8	7.0	1830121	4	US-09-557-884-1	Sequence 1, Appli
	9	31.8	7.0	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c	10	31.4	6.9	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	11	31	6.8	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	12	30.6	6.7	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	13	30.6	6.7	1350	4	US-09-328-352-3386	Sequence 3386, Ap
c	14	30.2	6.6	1053	4	US-09-543-681A-408	Sequence 408, App
	15	30.2	6.6	1072	4	US-08-956-171E-435	Sequence 435, App
	16	30	6.6	1454	2	US-08-657-392-1	Sequence 1, Appli
	17	30	6.6	1454	5	PCT-US94-02539-1	Sequence 1, Appli
	18	30	6.6	1455	2	US-08-657-392-26	Sequence 26, Appl
	19	30	6.6	1455	5	PCT-US94-02539-26	Sequence 26, Appl
	20	30	6.6	9072	4	US-08-956-171E-45	Sequence 45, Appl
	21	30	6.6	92407	4	US-09-596-002-36	Sequence 36, Appl
	22	29.8	6.5	640681	4	US-09-790-988-1	Sequence 1, Appli
c	23	29.6	6.5	474	4	US-09-621-976-18033	Sequence 18033, A
c	24	29.6	6.5	553	4	US-09-621-976-15491	Sequence 15491, A
	25	29.6	6.5	1588	1	US-08-446-908-3	Sequence 3, Appli
	26	29.6	6.5	1588	1	US-08-231-205A-3	Sequence 3, Appli
	27	29.6	6.5	1588	2	US-08-871-161-3	Sequence 3, Appli
	28	29.6	6.5	1589	4	US-09-023-655-1132	Sequence 1132, Ap
c	29	29.4	6.4	141	4	US-08-956-171E-2326	Sequence 2326, Ap
c	30	29.4	6.4	501	4	US-08-956-171E-654	Sequence 654, App
	31	29.4	6.4	1440	4	US-09-328-352-3130	Sequence 3130, Ap
	32	29.4	6.4	1710	4	US-09-543-681A-934	Sequence 934, App
	33	29.4	6.4	7253	4	US-09-268-347-35	Sequence 35, Appl
	34	29.4	6.4	9793	1	US-08-470-202-56	Sequence 56, Appl
	35	29.4	6.4	9793	1	US-08-471-770-56	Sequence 56, Appl
	36	29.4	6.4	9793	2	US-08-468-059-56	Sequence 56, Appl
	37	29.4	6.4	9793	3	US-09-109-916-56	Sequence 56, Appl
	38	29.4	6.4	9793	4	US-09-886-156-56	Sequence 56, Appl
	39	29.4	6.4	9793	4	US-09-886-149-56	Sequence 56, Appl
	40	29.4	6.4	9793	4	US-09-886-150-56	Sequence 56, Appl
	41	29.4	6.4	9793	4	US-09-886-159-56	Sequence 56, Appl
	42	29.2	6.4	1234	4	US-09-620-312D-953	Sequence 953, App
c	43	29.2	6.4	20284	4	US-09-526-193A-21	Sequence 21, Appl
	44	29	6.4	1287	4	US-09-461-325-18	Sequence 18, Appl
	45	29	6.4	1287	4	US-10-012-542-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 9.0%; Score 41; DB 1; Length 7218;

Best Local Similarity 3.0%; Pred. No. 0.0018;
Matches 11; Conservative 200; Mismatches 150; Indels 0; Gaps 0;

```
Qy          63 GACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTC 122
Db        1454 GAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395

Qy         123 TTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACG 182
Db        1394 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1335

Qy         183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAA 242
Db        1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1275

Qy         243 AGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGA 302
Db        1274 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1215

Qy         303 TGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGG 362
Db        1214 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155

Qy         363 AAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACA 422
Db        1154 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095

Qy         423 G 423
Db        1094 R 1094
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 7.5%; Score 34; DB 1; Length 361;
Best Local Similarity 56.1%; Pred. No. 0.091;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy         272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
Db         140 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 199

Qy         332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
Db         200 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 253
```

RESULT 3
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 7.5%; Score 34; DB 1; Length 456;
Best Local Similarity 56.1%; Pred. No. 0.1;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy         272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
```

```
Db        203 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 262

Qy        332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              || |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316


RESULT 4
US-09-328-352-2823
; Sequence 2823, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2823
;   LENGTH: 1443
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-2823

  Query Match             7.1%;  Score 32.2;  DB 4;  Length 1443;
  Best Local Similarity   46.9%;  Pred. No. 0.74;
  Matches  100;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||   ||| ||| | |||  |  | |      | |        |   ||||  || ||  ||
Db        481 ATATTAAAAAAACTATTAAATCAGTTACATGCTAATAATTATTTCCCTGTCATGGCTGCC 540

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              | | | | | | ||||         | |  || | | | | || |  |  || | |
Db        541 GAGCTAGAGTTTTATCTTTTCTCACCTCAGCATCAATCCGAGATATGTTCAGAAAACCAA 600

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              | ||    ||   | ||| | | || |  |||    |    |||| |   ||  | ||| |
Db        601 TGTTTTGATATAGATGCGCCAAATAATTATCAGCAAGTTCTGGATGAAGTTGAAAAAGCA 660

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATT 213
              ||    |  |||||    | |  ||| ||||
Db        661 GCCCTACTACAATCAATTGAAATTACCGCTATT 693


RESULT 5
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
;   GENERAL INFORMATION:
;        APPLICANT: Fleischmann et al.
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/557,884
;             FILING DATE: 25-Apr-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/476,102
;             FILING DATE: JUN-5-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Michelle S. Marks
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEFAX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             7.1%;  Score 32.2;  DB 4;  Length 1830121;
  Best Local Similarity   46.3%;  Pred. No. 25;
  Matches  106;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy          7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
              | ||| |  |  |  | |||   || |||| |||   |       ||   |||    |
Db     385834 ATCAACTATTGTTAAAGGATTCGGAAAATAATGGACTCTTTCTCGACATCAATTTGGGTT 385775

Qy         67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
              ||| |||      | |||||   |  ||  |  | | |  | | |    |  | ||  |||
Db     385774 AATGATGTAACTGTTCGTTACAACAATGGTCATACTGCTATTCACAATATGACGTTCTCA 385715

Qy        127 TTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAG 186
              | |||   |   | |||  ||    ||| |||  | ||   |||| ||
Db     385714 CTCAATAGTGGAACTATTTGTGCGTTAGTTGGCGTGAATGGAAGTGGCAAGTCCACATTA 385655

Qy        187 GAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATC 235
                  |  | | |   |  || |  ||||| |||| | || ||| |
Db     385654 TTTAAAAGCATAATGGGGCTTGTAAAGCCACAACAAGGCGAAATTAAAC 385606


RESULT 6
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;        APPLICANT: Robert D. Fleischmann
;                   Mark D. Adams
;                   Owen White
;                   Hamilton O. Smith
;                   J. Craig Venter
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville,
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/643,990A
;             FILING DATE: 23-Aug-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/487,429
;             FILING DATE: 1995-06-07
;             APPLICATION NUMBER: 08/426,787
```

; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 7.1%; Score 32.2; DB 4; Length 1830121;
Best Local Similarity 46.3%; Pred. No. 25;
Matches 106; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
Db 385834 ATCAACTATTGTTAAAGGATTCGGAAAATAATGGACTCTTTCTCGACATCAATTTGGGTT 385775

Qy 67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
Db 385774 AATGATGTAACTGTTCGTTACAACAATGGTCATACTGCTATTCACAATATGACGTTCTCA 385715

Qy 127 TTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAG 186
Db 385714 CTCAATAGTGGAACTATTTGTGCGTTAGTTGGCGTGAATGGAAGTGGCAAGTCCACATTA 385655

Qy 187 GAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATC 235
Db 385654 TTTAAAAGCATAATGGGGCTTGTAAAGCCACAACAAGGCGAAATTAAAC 385606

RESULT 7
US-09-134-000C-1323/c
; Sequence 1323, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1323
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1323

Query Match 7.0%; Score 32; DB 4; Length 651;
Best Local Similarity 55.3%; Pred. No. 0.58;
Matches 83; Conservative 0; Mismatches 65; Indels 2; Gaps 1;

Qy 217 CAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTAT 276
Db 178 CAAGAACTACGATAAAAGCCAGTAAAATACTCCCTGGTAGGCCTAAATAGTAATTGGCAA 119

Qy 277 ATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGT 336
Db 118 TTTGGGCTAA--GGCAATCCCGCCGTTTTCACTGATTGGAAAAGCCCCGATACTCATTGT 61

Qy 337 GCAGCTATTATCCAGAAAGGTTCTGGAAAT 366

Db 60 TCCCATATAAGCAAGTAAGGTATAGATAAT 31

RESULT 8
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 7.0%; Score 31.8; DB 4; Length 1830121;
Best Local Similarity 52.7%; Pred. No. 34;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
Db 617140 AACACCTTGTAAATTATGGTGTTCATCATAATAGGTTGGAATGCCACCGCCCCCCGCACA 617199

Qy 67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
Db 617200 AATTACGATACTGCCTTTTTCAAGCAACCATTTGACCGGACGAATTTCAAAAATACGTTT 617259

Qy 127 TTTAATCAGGC 137
Db 617260 TGGTAATGGGC 617270

RESULT 9
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White

```
;                  Hamilton O. Smith
;                  J. Craig Venter
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville,
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/643,990A
;            FILING DATE: 23-Aug-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/487,429
;            FILING DATE: 1995-06-07
;            APPLICATION NUMBER: 08/426,787
;            FILING DATE: 1995-04-21
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Kenley K. Hoover
;            REGISTRATION NUMBER: 40,302
;            REFERENCE/DOCKET NUMBER: PB186P1C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-610-5790
;            TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match             7.0%;  Score 31.8;  DB 4;  Length 1830121;
  Best Local Similarity   52.7%;  Pred. No. 34;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
              |||| |||| | ||||| |||| |  ||| | |||  |  ||  |  ||  ||  |||
Db     617140 AACACCTTGTAAATTATGGTGTTCATCATAATAGGTTGGAATGCCACCGCCCCCCGCACA 617199

Qy         67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
              ||||| |||   ||   ||   |  |    ||||      |   | || |  ||  | |
Db     617200 AATTACGATACTGCCTTTTTCAAGCAACCATTTGACCGGACGAATTTCAAAAATACGTTT 617259

Qy        127 TTTAATCAGGC 137
              |   |   |||
Db     617260 TGGTAATGGGC 617270


RESULT 10
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849

; SEQ ID NO 1
;  LENGTH: 1230025
;  TYPE: DNA
;  ORGANISM: Chlamydia pneumoniae
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(15000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (15001)..(30000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (30001)..(45000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (45001)..(60000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (60001)..(75000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (75001)..(90000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (90001)..(105000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (105001)..(120000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (120001)..(135000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (135001)..(150000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (150001)..(165000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (165001)..(180000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (180001)..(195000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (195001)..(210000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (210001)..(225000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (225001)..(240000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (240001)..(255000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (255001)..(270000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (270001)..(285000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (285001)..(300000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (300001)..(315000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (315001)..(330000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (330001)..(345000)
```

; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (720001)..(735000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (735001)..(750000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (750001)..(765000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (765001)..(780000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (780001)..(795000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (795001)..(810000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (810001)..(825000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

Query Match 6.9%; Score 31.4; DB 4; Length 1230025;
Best Local Similarity 56.2%; Pred. No. 38;
Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy        68 ATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCAT 127
             ||| |  |||| | |  ||||| | ||||  | |  | |  |||| ||    | ||||  |
Db    739726 ATCAAAATCTTGATTTTTCAGGAAATAAAATAGGCGAAGTTGAAGTAGCAGACTCTTTGT 739667

Qy       128 TTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTG 172
             ||     | | ||  |  |     ||||  |    || ||||  ||| ||
Db    739666 TTGCAGATGAGGGCGATGGCCTTCAATTGATCAAGGATTATATTG 739622
```

RESULT 11
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA

; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (300001)..(315000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t

```
;   NAME/KEY: misc_feature
;   LOCATION: (720001)..(735000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (735001)..(750000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (750001)..(765000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (765001)..(780000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (780001)..(795000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (795001)..(810000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (810001)..(825000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (825001)..(840000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (840001)..(855000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (855001)..(870000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (870001)..(885000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (885001)..(900000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (900001)..(915000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature

  Query Match             6.8%;  Score 31;  DB 4;  Length 1230025;
  Best Local Similarity   56.3%;  Pred. No. 52;
  Matches   58;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        344 TTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAG 403
              | |||    | | | | | || | | |||  ||||| | |      | | |    | |||||
Db    1044632 TGATCAGAAGAAATGCAGGCAGTGAGGTAAATATGAATCGTAGAGACATGGTAATAACAG 1044691

Qy        404 CAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCAC 446
              | || |||||| |       | || | |||||  ||  ||||||
Db    1044692 CTGTCGTAGTGAATGCTATATTGCTTGTGGCTCTTTTCGTCAC 1044734


RESULT 12
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 832;
  Best Local Similarity   11.1%;  Pred. No. 1.9;
  Matches   21;  Conservative   92;  Mismatches   76;  Indels    0;  Gaps    0;

Qy         10 AAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACAAAT 69
              ::::::: :::::: :| :: ::|::  ::::  |:  |::: ::|::::   |: :
Db         40 WRRKKKKAWWKYWKTWTWWYWRYAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYM 99

Qy         70 TATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTT 129
              :: :| ||  :::  :  :::: : :: :  : |     : :|:: :    : |: :::|
Db        100 AMWTRTWTGYAYYRSMMYWWRYRCWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKT 159

Qy        130 AATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAA 189
              : | ::  :: :|  ::   :  :  | | :  |::: :   :|:  ::| :: :  |
Db        160 YYWAATRYWWMMCWTKRWRASWWYCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGA 219

Qy        190 GGATCAAAA 198
              :::::::::
Db        220 MSWKYMWRM 228


RESULT 13
US-09-328-352-3386/c
; Sequence 3386, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3386
;   LENGTH: 1350
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-3386

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 1350;
  Best Local Similarity   65.2%;  Pred. No. 2.5;
  Matches   45;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        309 TATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAA 368
              |||||   |||||  ||   |  || |||   |||| ||||  ||| ||| | ||||||
Db        586 TATATGAAAAAGCTGTTTGTGAGCTAATGCCAGTATTTTCCATTAAGCTTCGGAAAATAA 527

Qy        369 GGCCAATAT 377
              |||| | |
Db        526 TGCCACTTT 518


RESULT 14
US-09-543-681A-408/c
; Sequence 408, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 408
;   LENGTH: 1053
;   TYPE: DNA
```

; ORGANISM: Proteus mirabilis
US-09-543-681A-408

Query Match 6.6%; Score 30.2; DB 4; Length 1053;
Best Local Similarity 62.7%; Pred. No. 3;
Matches 47; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
Qy        322 GCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACC 381
              || | |  ||| | | ||    ||||| || || |     ||||||||||| |   |||
Db        366 GCCTTCATTAACATTACAAGAGTTATCAAGGAAAGGATCAGAAATAAGGCCTTTCCAACC 307

Qy        382 CAGTACGGTACGCAG 396
               |  |||||||| |
Db        306 AATAACGGTACGCGG 292
```

RESULT 15
US-08-956-171E-435
; Sequence 435, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 435:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1072 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 435:
US-08-956-171E-435

Query Match 6.6%; Score 30.2; DB 4; Length 1072;
Best Local Similarity 49.3%; Pred. No. 3;
Matches 68; Conservative 3; Mismatches 67; Indels 0; Gaps 0;

```
Qy         98 TTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCG 157
              ||  |||||: ||   |: |||      |    || |:|||    | || | |  ||  |
Db          6 TTTTGGTAARGGAGAAAWTCAATATGAAACAGTATAARGCGTATTTAATCGATTTAGATG 65

Qy        158 GCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCAC 217
              ||||     |||  |  | ||         || ||| ||| || |||   || |  || |
Db         66 GCACAATGTATATGGGAACAGATGAGATTGATGGAGCAANACAATTCATCGATTATTTAA 125

Qy        218 AAGAAGGAGGAAATAATC 235
              | | |  ||| | |  ||
Db        126 ATGTAAAAGGCATTCCTC 143
```

Search completed: March 18, 2004, 20:19:39
Job time : 65.4971 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	35	7.7	738	15	US-10-369-435-3	Sequence 3, Appli
c 2	34.4	7.5	18540	12	US-10-231-416-3	Sequence 3, Appli
c 3	34.4	7.5	18540	14	US-10-231-426-3	Sequence 3, Appli
4	34	7.5	529	15	US-10-027-632-241642	Sequence 241642,
c 5	33.8	7.4	203654	9	US-09-820-905-3	Sequence 3, Appli
c 6	33.4	7.3	598	15	US-10-027-632-206661	Sequence 206661,
c 7	33.4	7.3	598	15	US-10-027-632-206662	Sequence 206662,
c 8	33.4	7.3	598	15	US-10-027-632-206663	Sequence 206663,
c 9	33.4	7.3	660	15	US-10-369-493-29323	Sequence 29323, A
10	33.2	7.3	535	12	US-10-449-857A-18	Sequence 18, Appl
11	33.2	7.3	1444	12	US-10-425-114-3093	Sequence 3093, Ap
12	32.8	7.2	2649	15	US-10-094-749-1500	Sequence 1500, Ap
13	32.4	7.1	1914	9	US-09-815-242-4167	Sequence 4167, Ap
14	32.4	7.1	1929	9	US-09-815-242-8045	Sequence 8045, Ap
15	32.2	7.1	1163020	15	US-10-398-221-10	Sequence 10, Appl
c 16	32.2	7.1	1830121	14	US-10-329-960-1	Sequence 1, Appli
c 17	32.2	7.1	1830121	15	US-10-329-670-1	Sequence 1, Appli
18	32.2	7.1	3011208	15	US-10-398-221-2058	Sequence 2058, Ap
19	32	7.0	849	15	US-10-027-632-130879	Sequence 130879,
20	32	7.0	849	15	US-10-027-632-130880	Sequence 130880,
21	32	7.0	849	15	US-10-027-632-130881	Sequence 130881,
c 22	32	7.0	1353	12	US-10-282-122A-21063	Sequence 21063, A
c 23	32	7.0	1356	9	US-09-815-242-6450	Sequence 6450, Ap
24	32	7.0	1539	12	US-10-282-122A-25267	Sequence 25267, A
c 25	32	7.0	2410	12	US-10-424-599-51912	Sequence 51912, A
c 26	32	7.0	7664	9	US-09-070-927A-375	Sequence 375, App
c 27	31.8	7.0	583	15	US-10-027-632-14367	Sequence 14367, A
c 28	31.8	7.0	933	9	US-09-815-242-6967	Sequence 6967, Ap
c 29	31.8	7.0	933	12	US-10-282-122A-22047	Sequence 22047, A
30	31.8	7.0	1365	9	US-09-815-242-4243	Sequence 4243, Ap
31	31.8	7.0	1365	9	US-09-815-242-8196	Sequence 8196, Ap
c 32	31.8	7.0	2209	9	US-09-925-301-494	Sequence 494, App
c 33	31.8	7.0	5599	12	US-10-363-616-89	Sequence 89, Appl
34	31.8	7.0	1830121	14	US-10-329-960-1	Sequence 1, Appli
35	31.8	7.0	1830121	15	US-10-329-670-1	Sequence 1, Appli
36	31.6	6.9	480	9	US-09-815-242-4053	Sequence 4053, Ap
37	31.6	6.9	480	12	US-10-282-122A-7419	Sequence 7419, Ap
c 38	31.6	6.9	1251	15	US-10-369-493-44921	Sequence 44921, A
39	31.6	6.9	465237	9	US-09-933-267A-1	Sequence 1, Appli
c 40	31.4	6.9	472	9	US-09-864-761-2780	Sequence 2780, Ap
c 41	31.4	6.9	664	15	US-10-027-632-135259	Sequence 135259,
c 42	31.4	6.9	664	15	US-10-027-632-135260	Sequence 135260,
c 43	31.4	6.9	664	15	US-10-027-632-152059	Sequence 152059,
c 44	31.4	6.9	808	15	US-10-027-632-160710	Sequence 160710,
c 45	31.4	6.9	808	15	US-10-027-632-160711	Sequence 160711,

ALIGNMENTS

RESULT 1
US-10-369-435-3
; Sequence 3, Application US/10369435
; Publication No. US20040002440A1
; GENERAL INFORMATION:
; APPLICANT: Mathews, Sarah
; APPLICANT: Timms, Peter
; TITLE OF INVENTION: No. US20040002440A1el Diagnostic Agents and Uses Therefor
; FILE REFERENCE: 10338-15US (2615070/VPA)
; CURRENT APPLICATION NUMBER: US/10/369,435
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: AU PQ9540/00
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/AU01/01021
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(738)
US-10-369-435-3

Query Match 7.7%; Score 35; DB 15; Length 738;
Best Local Similarity 57.9%; Pred. No. 1.9;
Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy        340 GCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAA 399
              ||| ||   |||| |  | | || | | |||  ||||| |  |     |  | |  | ||
Db          1 GCTGTTGATCAGAGAAATGCAGGCAGTGAGGTAAATATGAATCGTAGAGACATGGTAATA 60

Qy        400 ACAGCAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCAC 446
              ||||| || |||||| |       | || | ||||| ||  ||||||
Db         61 ACAGCTGTCGTAGTGAATGCTATATTGCTTGTGGCTCTTTTCGTCAC 107
```

RESULT 2
US-10-231-416-3/c
; Sequence 3, Application US/10231416
; Publication No. US20040043610A1
; GENERAL INFORMATION:
; APPLICANT: Bodian, Dale L.
; APPLICANT: Labow, Mark
; APPLICANT: Mickanin, Craig S.
; TITLE OF INVENTION: Tumor Necrosis Factor Related Protein-1
; TITLE OF INVENTION: Gene and Protein
; FILE REFERENCE: 4-32139A
; CURRENT APPLICATION NUMBER: US/10/231,416
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/317,151
; PRIOR FILING DATE: 2002-09-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 18540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-231-416-3

Query Match 7.5%; Score 34.4; DB 12; Length 18540;
Best Local Similarity 52.0%; Pred. No. 16;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        86 CAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTA 145
             ||||  |||||  |     | |  |    || || | |  |  |    | ||  |   |||
Db      5513 CAGATGGTAATTATTAAAAATAAAATGACAGAAATTTTAAAAATTCACTGGATGGGCTTA 5454

Qy       146 TTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGT 205
             | ||  ||  |  || ||| ||||   ||| | ||   ||||  || ||| | |||
Db      5453 GTAGTAGAGAAGAGACAGATGATAGAATCAGTGAACTTGAGGACAGAACAACAGAATTTA 5394

Qy       206 CCGTTATTTCACAAGAAGGAGGAAAATAA 233
                 | ||||  | ||||  |||| |
Db      5393 TGCAATATGAACAACAGGGAGAAAATGA 5366
```

RESULT 3
US-10-231-426-3/c
; Sequence 3, Application US/10231426
; Publication No. US20030059889A1
; GENERAL INFORMATION:
; APPLICANT: Bodian, Dale L.
; APPLICANT: Labow, Mark
; APPLICANT: Mickanin, Craig S.
; TITLE OF INVENTION: Tumor Necrosis Factor Related Protein-1
; TITLE OF INVENTION: Gene and Protein
; FILE REFERENCE: 4-32139A
; CURRENT APPLICATION NUMBER: US/10/231,426
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/317,151
; PRIOR FILING DATE: 2002-09-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 18540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-231-426-3

Query Match 7.5%; Score 34.4; DB 14; Length 18540;
Best Local Similarity 52.0%; Pred. No. 16;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        86 CAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTA 145
             ||||  |||||  |     | |  |    || || | |  |  |    | ||  |   |||
Db      5513 CAGATGGTAATTATTAAAAATAAAATGACAGAAATTTTAAAAATTCACTGGATGGGCTTA 5454

Qy       146 TTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGT 205
             | ||  ||  |  || ||| ||||   ||| | ||   ||||  || ||| | |||
Db      5453 GTAGTAGAGAAGAGACAGATGATAGAATCAGTGAACTTGAGGACAGAACAACAGAATTTA 5394

Qy       206 CCGTTATTTCACAAGAAGGAGGAAAATAA 233
                 | ||||  | ||||  |||| |
Db      5393 TGCAATATGAACAACAGGGAGAAAATGA 5366
```

RESULT 4
US-10-027-632-241642
; Sequence 241642, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 241642
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(529)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-241642

Query Match 7.5%; Score 34; DB 15; Length 529;
Best Local Similarity 54.9%; Pred. No. 3.4;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy       132 TCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGG 191
             | | | | ||| ||   ||   |    || | ||| |   |||   || | |||| |
Db       132 TGATGGGCCCAGTAGAAGTGTGGATATGACAGGGAATGGGAGCAGTTCACTCGAGGATAG 191

Qy       192 ATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCA 251
             ||||| |  | |  ||  | ||  |||| | |||| |||||  | |  |||  || ||
Db       192 ATCAATAAAACTTACCCATCTTGAACAACAGGGAGAAAATAGACTGAAAAAAAAAGAGCA 251

Qy       252 GG 253
               |
Db       252 AG 253
```

RESULT 5
US-09-820-905-3/c
; Sequence 3, Application US/09820905
; Patent No. US20020142938A1
; GENERAL INFORMATION:
; APPLICANT: Yan, Chunhua
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001199
; CURRENT APPLICATION NUMBER: US/09/820,905

```
;  CURRENT FILING DATE:  2001-03-30
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 203654
;   TYPE: DNA
;   ORGANISM: HUMAN
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(203654)
;   OTHER INFORMATION: n = A,T,C or G
US-09-820-905-3

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 203654;
  Best Local Similarity   50.3%;  Pred. No. 86;
  Matches   83;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy          5 AAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGA 64
              | ||  | || || ||| |  |  |     |||||    |  ||||||    |||   |
Db      71389 AGAAATATTTTTTCTTTTTTGTAGTATGTTTACTGAAAACTACTGGGAACAAAACATGGT 71330

Qy         65 CAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTT 124
              ||| | |      ||  || | | ||| |||  |||    | ||| | || || | |    |
Db      71329 CAACTGAACTAATGCTTCCTCCTGAAAATATGTTTTTAGAAAAAGTATAAACCCACAAAT 71270

Qy        125 CATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATA 169
                |  | ||||  |||     |  | ||  |  ||  | ||||
Db      71269 ACTAAGAACAGGAAACCAGATGAGAGTATGTGGAAACATAGAATA 71225


RESULT 6
US-10-027-632-206661/c
; Sequence 206661, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 206661
;   LENGTH: 598
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-206661

  Query Match             7.3%;  Score 33.4;  DB 15;  Length 598;
  Best Local Similarity   51.7%;  Pred. No. 5.6;
  Matches   76;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        200 TATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGA 259
              |||| | | ||| |||| |   |    | || |||  | | |||       ||  |
Db        563 TATTATTCTTTAATTCATAGAGATCTTGTAAGAATTAGCCTAAAACAATTAAGATACATT 504

Qy        260 ATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAA 319
                ||  ||  | || | ||||  |||   |   ||||| ||  ||   || || | | ||
Db        503 TTTTAAATGTGGCATTTATTAGGCAGTTGATAAATGCTAATTATTATAGCATTTAGTAAT 444

Qy        320 GCGCTTACGGTAATAGTGCAGCTATTA 346
                ||  | |   ||  | || || | | |
Db        443 GCTTTAATAATAGCATTGGAGGTGTGA 417


RESULT 7
US-10-027-632-206662/c
; Sequence 206662, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 206662
;   LENGTH: 598
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-206662

  Query Match             7.3%;  Score 33.4;  DB 15;  Length 598;
  Best Local Similarity   51.7%;  Pred. No. 5.6;
  Matches   76;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        200 TATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGA 259
              |||| | | ||| |||| |   |    | || |||  | | |||       ||  |
Db        563 TATTATTCTTTAATTCATAGAGATCTTGTAAGAATTAGCCTAAAACAATTAAGATACATT 504

Qy        260 ATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAA 319
                ||  ||  | || | ||||  |||   |   ||||| ||  ||   || || | | ||
Db        503 TTTTAAATGTGGCATTTATTAGGCAGTTGATAAATGCTAATTATTATAGCATTTAGTAAT 444

Qy        320 GCGCTTACGGTAATAGTGCAGCTATTA 346
                ||  | |   ||  | || || | | |
Db        443 GCTTTAATAATAGCATTGGAGGTGTGA 417


RESULT 8
US-10-027-632-206663/c
; Sequence 206663, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
```

; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 206663
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-206663

Query Match 7.3%; Score 33.4; DB 15; Length 598;
Best Local Similarity 51.7%; Pred. No. 5.6;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 200 TATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGA 259
Db 563 TATTATTCTTTAATTCATAGAGATCTTGTAAGAATTAGCCTAAAACAATTAAGATACATT 504
Qy 260 ATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAA 319
Db 503 TTTTAAATGTGGCATTTATTAGGCAGTTGATAAATGCTAATTATTATAGCATTTAGTAAT 444
Qy 320 GCGCTTACGGTAATAGTGCAGCTATTA 346
Db 443 GCTTTAATAATAGCATTGGAGGTGTGA 417

RESULT 9
US-10-369-493-29323/c
; Sequence 29323, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 29323
; LENGTH: 660
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-10-369-493-29323

Query Match 7.3%; Score 33.4; DB 15; Length 660;
Best Local Similarity 55.7%; Pred. No. 5.9;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 250 CAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGT 309
Db 339 CAATCCGGCAATTTGCCGATTTGCCAGATTTGCCGAACAGCAATTGCCGCCCATTCCTGA 280
Qy 310 ATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAA 364
Db 279 ATTATTTAGTGCATCTACAGTAATATCTTATAGAGTATCCATACAGGCTCTGGAA 225

RESULT 10
US-10-449-857A-18

; Sequence 18, Application US/10449857A
; Publication No. US20040043931A1
; GENERAL INFORMATION:
; APPLICANT: Hershberg, Robert M.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF INFLAMMATORY BOWEL DISEASE
; FILE REFERENCE: 584C1
; CURRENT APPLICATION NUMBER: US/10/449,857A
; CURRENT FILING DATE: 2003-05-30
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 18
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Unknown Bacterium
US-10-449-857A-18

Query Match 7.3%; Score 33.2; DB 12; Length 535;
Best Local Similarity 51.3%; Pred. No. 6.1;
Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 248 ACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCA 307
Db 14 ACCAGACTGCAAATCGTCACTATGCACACGTTGACTGCCCGGGCCACGCCGACTATGTGA 73
Qy 308 GTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATA 367
Db 74 AGAACATGGTAACTGGTGCAGCTCAGATGGACGGTGCAATCCTTGTTTGTGCTGCAACTG 133
Qy 368 AGGCCAATATTACCCAGTACGGTACGCAGA 397
Db 134 ACGGTCCTATGCCTCAGACACGTGAGCACA 163

RESULT 11
US-10-425-114-3093
; Sequence 3093, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3093
; LENGTH: 1444
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700240986_FLI
US-10-425-114-3093

Query Match 7.3%; Score 33.2; DB 12; Length 1444;
Best Local Similarity 55.1%; Pred. No. 10;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 249 CCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAG 308
Db 207 CCATGGAGTTGATAATCACCCACCTTACGTCAAGAAACAAGTGGTGGAAAAGTTTCCCTG 266
Qy 309 TATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAAT 366

```
                 | || | || ||| | || ||||| || || | ||||| ||| || || || |
Db        267 CTCAGCGTACAGTGCTCCCTGTTCTAGTGAAGATAATTTCCAGGAAGCTTGTGCAATT 324
```

RESULT 12
US-10-094-749-1500
; Sequence 1500, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1500
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1500

Query Match 7.2%; Score 32.8; DB 15; Length 2649;
Best Local Similarity 48.4%; Pred. No. 19;
Matches 91; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy        208 GTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAAC 267
              ||| ||   | || |||||  |   |    |  | ||   ||  | ||  |  | |  ||
Db        807 GTTCTTGATCTAGCAGGAGTGATACAAAACGTCATAGATCACGTGCCAAAGCTTCACTAC 866

Qy        268 TTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTAC 327
              | |    | | ||  |  | |   || ||   || | ||   |   |   | | | |
Db        867 CTGGTTGCTGTGGATGAGTCTGTTCATACCCTGGAAGACATAGTCAAGTGTATCAGTAAA 926

Qy        328 GGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTAC 387
              || | |  | |  |  |||||||||    |  || | || ||  |  | ||| || ||
Db        927 AATACTGGCCCTGGGAAAATCCAGAAAATACCCAGAGAAAATGCATACCTAACCAAGGAC 986

Qy        388 GGTACGCA 395
              |||||
Db        987 TTAACGCA 994
```

RESULT 13
US-09-815-242-4167
; Sequence 4167, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4167
; LENGTH: 1914
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4167

Query Match 7.1%; Score 32.4; DB 9; Length 1914;
Best Local Similarity 51.4%; Pred. No. 21;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        263 ATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCG 322
              | || || |    |  | | |||||  |  | | ||||   |  |||||||  |
Db       1001 AAAATTTAGAAATTGGTTATCAAACTGCAATTACCAAACCTATCAGTATAGAGGTCTCTA 1060

Qy        323 CTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCC 382
              || | |||  |||   |||   |  || |||  ||||||   | ||   ||||
Db       1061 AAGGCGATCATATAGCAATCATTGGGCCAAATGGTATTGGAAAATCGACCTTAATTAAAA 1120

Qy        383 AGTACGGTACGCAGAAAACAGCAGTT 408
              | ||  ||  ||| |||| ||
Db       1121 CTATTGCTAATCAACAAAAAGCACTT 1146
```

RESULT 14
US-09-815-242-8045
; Sequence 8045, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

```
  ;  PRIOR FILING DATE: 2000-10-23
  ;  PRIOR APPLICATION NUMBER: 60/253,625
  ;  PRIOR FILING DATE: 2000-11-27
  ;  PRIOR APPLICATION NUMBER: 60/257,931
  ;  PRIOR FILING DATE: 2000-12-22
  ;  PRIOR APPLICATION NUMBER: 60/269,308
  ;  PRIOR FILING DATE: 2001-02-16
  ;  NUMBER OF SEQ ID NOS: 14110
  ;  SOFTWARE: FastSEQ for Windows Version 4.0
  ; SEQ ID NO 8045
  ;    LENGTH: 1929
  ;    TYPE: DNA
  ;    ORGANISM: Staphylococcus aureus
  ;    FEATURE:
  ;    NAME/KEY: CDS
  ;    LOCATION: (1)...(1929)
 US-09-815-242-8045

    Query Match             7.1%;  Score 32.4;  DB 9;  Length 1929;
    Best Local Similarity   51.4%;  Pred. No. 21;
    Matches   75;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        263 ATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCG 322
              | || || |    |  | | |||||  |  | | ||||    |  |||||||  |
Db       1001 AAAATTTAGAAATTGGTTATCAAACTGCAATTACCAAACCTATCAGTATAGAGGTCTCTA 1060

Qy        323 CTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCC 382
              || | |||  |||  ||| .  |  || |||  ||||||   | ||  ||||
Db       1061 AAGGCGATCATATAGCAATCATTGGGCCAAATGGTATTGGAAAATCGACCTTAATTAAAA 1120

Qy        383 AGTACGGTACGCAGAAAACAGCAGTT 408
                | ||  ||  ||| |||| ||
Db       1121 CTATTGCTAATCAACAAAAAGCACTT 1146


RESULT 15
US-10-398-221-10
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;    LENGTH: 1163020
;    TYPE: DNA
;    ORGANISM: Listeria innocua
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)..(end)
;    OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

    Query Match             7.1%;  Score 32.2;  DB 15;  Length 1163020;
    Best Local Similarity   49.8%;  Pred. No. 6.2e+02;
    Matches  108;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        215 CACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGT 274
              |||||||   ||  || ||| | ||  || |||   |   ||  ||  ||   || | |
Db    1136649 CACAAGACATAGAGAAGAATTGTGCTGAATTCGTTAACTTAGATAAAAATCTATTAACTG 1136708

Qy        275 ATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATA 334
              |||||||  ||||||      |  |||    || ||  || ||   | |  |  |||
Db    1136709 ATATTGAAGTAACGGGGTGATTGTAAAATGAAGATAGATGTCAGAGAATTAAATGAGATA 1136768

Qy        335 GTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGC 394
              ||  ||| |  ||  || || | |   ||||  |     || |||      ||
Db    1136769 GT-TAGCGAGAATATAGTAATGCTTAAGAAAGAGTATGAAAATTTAAATATAGCAATGAA 1136827

Qy        395 AGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATAT 431
              | |  | || ||  |  | ||||||    | || |
Db    1136828 CGCTACAACCAATTTCACAGTAGAAACTCAGGAATTT 1136864


Search completed: March 19, 2004, 03:48:35
Job time : 260.734 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-2	Sequence 2, Appli
2	332.8	73.0	456	23	US-09-543-407-4	Sequence 4, Appli
3	295	64.7	561	17	US-09-252-691-171	Sequence 171, App
4	295	64.7	561	17	US-09-252-691C-171	Sequence 171, App
5	295	64.7	561	51	US-10-417-886-171	Sequence 171, App
6	74.4	16.3	100	45	US-10-146-492B-80	Sequence 80, Appl
7	66.4	14.6	100	45	US-10-146-492B-82	Sequence 82, Appl
8	64.8	14.2	100	45	US-10-146-492B-81	Sequence 81, Appl
9	44.2	9.7	438	31	US-09-739-449-3383	Sequence 3383, Ap
10	44.2	9.7	438	33	US-09-803-110-3383	Sequence 3383, Ap
c 11	44.2	9.7	29058	21	US-09-514-000-342	Sequence 342, App
12	44.2	9.7	187892	31	US-09-739-449-208	Sequence 208, App
13	44.2	9.7	187892	33	US-09-803-110-208	Sequence 208, App
c 14	41	9.0	7218	8	US-08-466-194-14	Sequence 14, Appl
15	37.6	8.2	456	23	US-09-543-407-13	Sequence 13, Appl
c 16	37.6	8.2	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
17	36.6	8.0	500	55	US-10-764-432-554	Sequence 554, App
18	36	7.9	1230230	20	US-09-438-185-1	Sequence 1, Appli
19	36	7.9	1230230	20	US-09-438-185A-1	Sequence 1, Appli
20	35.8	7.9	39	23	US-09-543-407-51	Sequence 51, Appl
21	35	7.7	738	50	US-10-369-435-3	Sequence 3, Appli
22	35	7.7	105325	1	PCT-US03-38193-1574	Sequence 1574, Ap
23	35	7.7	105325	54	US-10-723-860-1574	Sequence 1574, Ap
c 24	34.6	7.6	1362	104	US-60-475-871-2480	Sequence 2480, Ap
25	34.4	7.5	2006	46	US-10-160-619-101	Sequence 101, App
c 26	34.4	7.5	18540	47	US-10-231-416-3	Sequence 3, Appli
c 27	34.4	7.5	18540	47	US-10-231-426-3	Sequence 3, Appli
c 28	34.4	7.5	20537	39	US-09-949-016-12887	Sequence 12887, A
c 29	34.4	7.5	20538	39	US-09-949-016-17052	Sequence 17052, A
c 30	34.4	7.5	432287	39	US-09-947-911-180	Sequence 180, App
31	34.4	7.5	3227788	39	US-09-947-911-51	Sequence 51, Appl
32	34.2	7.5	456	23	US-09-543-407-25	Sequence 25, Appl
c 33	34.2	7.5	12757	36	US-09-897-516-2547	Sequence 2547, Ap
c 34	34.2	7.5	12757	36	US-09-897-516-2548	Sequence 2548, Ap
c 35	34.2	7.5	12757	36	US-09-897-516-2549	Sequence 2549, Ap
c 36	34.2	7.5	12757	36	US-09-897-516-2550	Sequence 2550, Ap
37	34.2	7.5	12757	36	US-09-897-516-2551	Sequence 2551, Ap
38	34.2	7.5	12757	36	US-09-897-516-2552	Sequence 2552, Ap
39	34.2	7.5	12757	36	US-09-897-516-2553	Sequence 2553, Ap
40	34.2	7.5	12757	36	US-09-897-516-2554	Sequence 2554, Ap
41	34.2	7.5	12757	36	US-09-897-516-2555	Sequence 2555, Ap
42	34.2	7.5	12757	36	US-09-897-516-2556	Sequence 2556, Ap
c 43	34.2	7.5	12757	36	US-09-897-516A-2547	Sequence 2547, Ap
c 44	34.2	7.5	12757	36	US-09-897-516A-2548	Sequence 2548, Ap
c 45	34.2	7.5	12757	36	US-09-897-516A-2549	Sequence 2549, Ap

ALIGNMENTS

```
RESULT 1
US-09-543-407-2
; Sequence 2, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
```

```
US-09-543-407-2

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 4.9e-130;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
```

```
RESULT 2
US-09-543-407-4
; Sequence 4, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-4

  Query Match           73.0%;  Score 332.8;  DB 23;  Length 456;
  Best Local Similarity  83.1%;  Pred. No. 7.2e-92;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  |||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
```

```
Db        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Db        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Db        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Db        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Db        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456

Db        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456


RESULT 3
US-09-252-691-171
; Sequence 171, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-171

  Query Match             64.7%;  Score 295;  DB 17;  Length 561;
  Best Local Similarity   78.0%;  Pred. No. 3.9e-80;
  Matches  355;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTA 455

Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 4
US-09-252-691C-171
; Sequence 171, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-171

  Query Match             64.7%;  Score 295;  DB 17;  Length 561;
  Best Local Similarity   78.0%;  Pred. No. 3.9e-80;
  Matches  355;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTA 455

Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 5
US-10-417-886-171
; Sequence 171, Application US/10417886
; GENERAL INFORMATION:
;  APPLICANT: Keith G. Weinstock et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.135
;  CURRENT APPLICATION NUMBER: US/10/417,886
;  CURRENT FILING DATE:  2003-04-17
```

```
;  PRIOR APPLICATION NUMBER: US/09/252,691C
;  PRIOR FILING DATE: 1999-02-18
;  PRIOR APPLICATION NUMBER: US 60/094,145
;  PRIOR FILING DATE: 1998-07-24
;  PRIOR APPLICATION NUMBER: US 60/074,787
;  PRIOR FILING DATE: 1998-02-18
;  NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 171
;    LENGTH: 561
;    TYPE: DNA
;    ORGANISM: Enterobacter cloacae
US-10-417-886-171

  Query Match                 64.7%;  Score 295;  DB 51;  Length 561;
  Best Local Similarity       78.0%;  Pred. No. 3.9e-80;
  Matches  355;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||  ||| |||||||||||  ||| ||||||||||||||||||  ||| ||| |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||     ||| ||| | ||  |||| ||||||||||||||||  ||||||||||||| ||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||  |||||||| ||||||||||||||||  ||  | |  |||| | |  ||| |
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||| || || ||| |  |||||||||||||  || || || ||||| || |||||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              || || || ||| |||||| |||| |||| ||||| || || || ||||| | ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||  |||||   ||| ||||||  || || |  ||  ||||||||||||| ||||||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGTCTG 465

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || || | ||| ||||| || |||| ||||||||||||||||||||||||||||| |||||
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTA 455
              |||||||| ||||||||||||||| |  ||||| ||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 6
US-10-146-492B-80
; Sequence 80, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 80
;    LENGTH: 100
;    TYPE: DNA
;    ORGANISM: Escherichia coli K-12 MG1655
;    FEATURE:
;    OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;    RELEVANT RESIDUES: 39-58
US-10-146-492B-80

  Query Match                 16.3%;  Score 74.4;  DB 45;  Length 100;
  Best Local Similarity       84.0%;  Pred. No. 4.7e-12;
  Matches   84;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        266 ACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTT 325
              || ||||| |||||||| ||  |||||| |||||||||||||||||| |||||| | ||||
Db          1 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 60

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAA 365
              | ||||||| |||     ||||||||||||||||||||| ||
Db         61 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 100


RESULT 7
US-10-146-492B-82
; Sequence 82, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 82
;    LENGTH: 100
;    TYPE: DNA
;    ORGANISM: Escherichia coli K-12 MG1655
;    FEATURE:
;    OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;    RELEVANT RESIDUES: 57-76
US-10-146-492B-82

  Query Match                 14.6%;  Score 66.4;  DB 45;  Length 100;
  Best Local Similarity       79.0%;  Pred. No. 1.4e-09;
  Matches   79;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        123 TTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACG 182
              ||||||||||||||| ||||| |||||||||  || ||  ||||||||     ||||
Db          1 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 60

Qy        183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAA 222
              |||| ||| |||||||| ||| | ||| || | ||||||
Db         61 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 100


RESULT 8
US-10-146-492B-81
; Sequence 81, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 81
;    LENGTH: 100
;    TYPE: DNA
;    ORGANISM: Escherichia coli K-12 MG1655
;    FEATURE:
;    OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;    RELEVANT RESIDUES: 38-57
US-10-146-492B-81

  Query Match                 14.2%;  Score 64.8;  DB 45;  Length 100;
  Best Local Similarity       78.0%;  Pred. No. 4.4e-09;
  Matches   78;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;
```

```
Qy        208 GTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAAC 267
              ||| || | |||||||| | | || ||||| || | |||||| |||| ||||||||
Db          1 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 60

Qy        268 TTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCA 307
              |||| |||||||| || |||||| ||||||||||||||
Db         61 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 100


RESULT 9
US-09-739-449-3383
; Sequence 3383, Application US/09739449
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)C
;  CURRENT APPLICATION NUMBER: US/09/739,449
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 3383
;   LENGTH: 438
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
US-09-739-449-3383

  Query Match             9.7%;  Score 44.2;  DB 31;  Length 438;
  Best Local Similarity   47.6%;  Pred. No. 0.017;
  Matches  130;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |     ||||||| |
Db         88 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 147

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
              |   |  | ||  | || ||  | || |||       | | |||||   ||   | |
Db        148 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 207

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| || ||  | | |  ||    ||  ||| ||| | |     |  | || | ||
Db        208 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 267

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
              || | ||    || |  ||   ||||  || ||     | |   |||   | ||||||
Db        268 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 327

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
              ||||  |     || |  || |  |  || |||
Db        328 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 360


RESULT 10
US-09-803-110-3383
; Sequence 3383, Application US/09803110
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)D
;  CURRENT APPLICATION NUMBER: US/09/803,110
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/739,449
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  PRIOR APPLICATION NUMBER: US 60/168,139
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 3383
;   LENGTH: 438
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
US-09-803-110-3383

  Query Match             9.7%;  Score 44.2;  DB 33;  Length 438;
  Best Local Similarity   47.6%;  Pred. No. 0.017;
  Matches  130;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |     ||||||| |
Db         88 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 147

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
              |   |  | ||  | || ||  | || |||       | | |||||   ||   | |
Db        148 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 207

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| || ||  | | |  ||    ||  ||| ||| | |     |  | || | ||
Db        208 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 267

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
              || | ||    || |  ||   ||||  || ||     | |   |||   | ||||||
Db        268 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 327

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
              ||||  |     || |  || |  |  || |||
Db        328 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 360


RESULT 11
US-09-514-000-342/c
; Sequence 342, Application US/09514000
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT: Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)B
;  CURRENT APPLICATION NUMBER: US/09/514,000
;  CURRENT FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 15034
; SEQ ID NO 342
;   LENGTH: 29058
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
US-09-514-000-342

  Query Match             9.7%;  Score 44.2;  DB 21;  Length 29058;
  Best Local Similarity   47.6%;  Pred. No. 0.073;
  Matches  130;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |     ||||||| |
Db      18682 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 18623

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
              |   |  | ||  | || ||  | || |||       | | |||||   ||   | |
Db      18622 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 18563

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| || ||  | | |  ||    ||  ||| ||| | |     |  | || | ||
Db      18562 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 18503

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
              || | ||    || |  ||   ||||  || ||     | |   |||   | ||||||
Db      18502 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 18443

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
              ||||  |     || |  || |  |  || |||
Db      18442 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 18410


RESULT 12
US-09-739-449-208
; Sequence 208, Application US/09739449
```

```
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)C
;  CURRENT APPLICATION NUMBER: US/09/739,449
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 208
;   LENGTH: 187892
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(187892)
;   OTHER INFORMATION: unsure at all n locations
US-09-739-449-208

  Query Match             9.7%;  Score 44.2;  DB 31;  Length 187892;
  Best Local Similarity   47.6%;  Pred. No. 0.14;
  Matches  130;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |      ||||||| |
Db     137108 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 137167

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
              |    |  |  ||  | || ||  | || |||       | | |||||   ||   | |
Db     137168 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 137227

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| || ||  | | |  ||     ||  ||| |||  | |       | || | ||
Db     137228 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 137287

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
              || | ||    || |  ||   ||||   || ||     | |   |||   | ||||||
Db     137288 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 137347

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
              ||||  |    || |  || |  |  || |||
Db     137348 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 137380


RESULT 13
US-09-803-110-208
; Sequence 208, Application US/09803110
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)D
;  CURRENT APPLICATION NUMBER: US/09/803,110
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/739,449
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  PRIOR APPLICATION NUMBER: US 60/168,139
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 208
;   LENGTH: 187892
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(187892)
;   OTHER INFORMATION: unsure at all n locations
US-09-803-110-208

  Query Match             9.7%;  Score 44.2;  DB 33;  Length 187892;
  Best Local Similarity   47.6%;  Pred. No. 0.14;
  Matches  130;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |      ||||||| |
Db     137108 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 137167

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
              |    |  |  ||  | || ||  | || |||       | | |||||   ||   | |
Db     137168 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 137227

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| || ||  | | |  ||     ||  ||| |||  | |       | || | ||
Db     137228 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 137287

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
              || | ||    || |  ||   ||||   || ||     | |   |||   | ||||||
Db     137288 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 137347

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
              ||||  |    || |  || |  |  || |||
Db     137348 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 137380


RESULT 14
US-08-466-194-14/c
; Sequence 14, Application US/08466194
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, NW, Suite 500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20007
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,194
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/232,463
;      FILING DATE:  22-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/935,313
;      FILING DATE:  26-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/201 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
```

US-08-466-194-14

Query Match 9.0%; Score 41; DB 8; Length 7218;
Best Local Similarity 3.0%; Pred. No. 0.44;
Matches 11; Conservative 200; Mismatches 150; Indels 0; Gaps 0;

```
Qy          63 GACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTC 122
               || | |  |  || |||  |:   :::: : ::    : :: ::: :::  ::: :::
Db        1454 GAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395

Qy         123 TTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACG 182
                :   ::  ::: ::  :  :  ::  :::  :: : ::: :: :: :  :::: : :
Db        1394 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1335

Qy         183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAA 242
                ::::::::  ::::  :  :   :  :    : :::::::::::: ::  ::: :::
Db        1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1275

Qy         243 AGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGA 302
               ::  ::  ::: :::::  : ::    : : : :  ::: ::: ::: :: :  :: ::
Db        1274 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1215

Qy         303 TGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGG 362
                :  :: : :  : :::: :   : :: :: :: : ::  :  :   :::::::    ::
Db        1214 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155

Qy         363 AAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACA 422
               ::: ::::  :: :  :   :: : :: : : :::::: :: ::  : :: : ::::: :
Db        1154 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095

Qy         423 G 423
               :
Db        1094 R 1094
```

RESULT 15
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match 8.2%; Score 37.6; DB 23; Length 456;
Best Local Similarity 51.2%; Pred. No. 2;
Matches 88; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy         272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
               ||   |||   || |  ||  |||  ||||  |||  | ||   ||  | ||  |   ||
Db         203 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 262

Qy         332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTA 391
                || |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db         263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACG 322

Qy         392 CGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGT 443
```

```
                   | | ||  | |   ||   ||     || |    ||  ||| ||  || ||
Db         323 CTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGT 374
```

Search completed: March 19, 2004, 03:01:39
Job time : 2655.77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	36.6	8.0	500	1	PCT-US04-02000-554	Sequence 554, App
c	2	33.8	7.4	401433	6	US-10-765-790-79	Sequence 79, Appl
	3	32.8	7.2	418	6	US-10-779-543-9321	Sequence 9321, Ap
c	4	32.4	7.1	201	6	US-10-767-471-8836	Sequence 8836, Ap
c	5	32.4	7.1	201	6	US-10-767-471-34493	Sequence 34493, A
c	6	32.4	7.1	1227	6	US-10-767-471-501	Sequence 501, App
c	7	32.4	7.1	76573	6	US-10-767-471-10718	Sequence 10718, A
	8	32	7.0	1310	6	US-10-767-795-2887	Sequence 2887, Ap
c	9	31.8	7.0	5495	6	US-10-775-169-342	Sequence 342, App
c	10	31.2	6.8	201	6	US-10-767-471-8837	Sequence 8837, Ap
c	11	31.2	6.8	201	6	US-10-767-471-34439	Sequence 34439, A
	12	31	6.8	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	13	30.8	6.8	309	6	US-10-257-047-43	Sequence 43, Appl
	14	30.8	6.8	309	6	US-10-257-047-46	Sequence 46, Appl
	15	30.8	6.8	3147	6	US-10-257-047-44	Sequence 44, Appl
	16	30.8	6.8	3147	6	US-10-257-047-45	Sequence 45, Appl
	17	30.8	6.8	3147	6	US-10-257-047-47	Sequence 47, Appl
	18	30.8	6.8	3147	6	US-10-257-047-48	Sequence 48, Appl
	19	29.8	6.5	592	6	US-10-767-701-25359	Sequence 25359, A
c	20	29.8	6.5	1035	6	US-10-779-543-7983	Sequence 7983, Ap
	21	29.8	6.5	2583	6	US-10-441-926-25	Sequence 25, Appl
	22	29.8	6.5	2583	6	US-10-441-926-27	Sequence 27, Appl
	23	29.8	6.5	2583	6	US-10-441-926-29	Sequence 29, Appl
	24	29.6	6.5	1084	6	US-10-487-578-48	Sequence 48, Appl
	25	29.6	6.5	1087	6	US-10-487-578-33	Sequence 33, Appl
	26	29.6	6.5	1217	6	US-10-487-578-46	Sequence 46, Appl
	27	29.6	6.5	1219	6	US-10-487-578-44	Sequence 44, Appl
	28	29.4	6.4	782	6	US-10-767-701-3762	Sequence 3762, Ap
c	29	29.4	6.4	188053	6	US-10-417-375A-125	Sequence 125, App
	30	29.2	6.4	344805	6	US-10-779-271-1	Sequence 1, Appli
	31	29.2	6.4	354592	6	US-10-765-790-70	Sequence 70, Appl
	32	29	6.4	1287	6	US-10-100-683-1253	Sequence 1253, Ap
c	33	29	6.4	1382	6	US-10-100-683-2753	Sequence 2753, Ap
c	34	29	6.4	113486	6	US-10-417-375A-18	Sequence 18, Appl
	35	29	6.4	156652	6	US-10-765-790-48	Sequence 48, Appl
	36	28.8	6.3	19959	6	US-10-767-471-10755	Sequence 10755, A
c	37	28.8	6.3	53118	6	US-10-767-471-10892	Sequence 10892, A
	38	28.6	6.3	346	6	US-10-779-543-12419	Sequence 12419, A
	39	28.6	6.3	675	5	US-09-741-873C-3	Sequence 3, Appli
	40	28.6	6.3	1743	6	US-10-417-884A-2080	Sequence 2080, Ap
c	41	28.6	6.3	3537	6	US-10-486-977-58	Sequence 58, Appl
	42	28.6	6.3	219715	6	US-10-417-375A-63	Sequence 63, Appl
	43	28.4	6.2	1149	6	US-10-771-241-139	Sequence 139, App
	44	28.4	6.2	2526	7	US-60-548-091-9	Sequence 9, Appli
	45	28.4	6.2	2665	7	US-60-548-091-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
PCT-US04-02000-554
; Sequence 554, Application PC/TUS0402000
; GENERAL INFORMATION:
; APPLICANT: Science Applications International Corporation
; TITLE OF INVENTION: Method and System for Identifying Biological Entities in
; TITLE OF INVENTION: Biological and Environmental Samples
; FILE REFERENCE: 36609-183264 (SAIC0087-PCT)
; CURRENT APPLICATION NUMBER: PCT/US04/02000
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US 60/441,745
; PRIOR FILING DATE: 2003-01-23
; PRIOR APPLICATION NUMBER: US 60/441,806
; PRIOR FILING DATE: 2003-01-23
; NUMBER OF SEQ ID NOS: 3242
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 554
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Clostridium perfringens
PCT-US04-02000-554

Query Match 8.0%; Score 36.6; DB 1; Length 500;
Best Local Similarity 44.4%; Pred. No. 0.053;
Matches 147; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

```
Qy          57 AACCGCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAG 116
               || | ||  || | | |||    |||||  | |   ||      |  ||||| |||  |
Db          68 AAACTCGTGAATTAAAGATGAAACTCGTGAAAAGGTTATGAAAGTTATAAATAAATATAA 127

Qy         117 CAAGTCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAG 176
               | |  || |   |  |  ||   | | ||   ||       ||  ||| || ||
Db         128 CTATACTCCTAGTGCTATTGCTAGAAGTCTTTCAAAAAGTATTACTAATACTATTGGAGT 187

Qy         177 AGTACGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCG 236
               ||   |  ||     |||| | ||| |  |       |  | |||||  || | ||
Db         188 TATAGTTCCTGATATTACAAATCCATTCTTTGGATCAATAATAAAAGGAATAAGTGATGT 247

Qy         237 GGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGC 296
               ||  ||||  |  |             |   | ||  | ||  |||   |  |  |  |
Db         248 TGCAGAAGTGCATAATCTTAATTTGATTCTTTGTGACTCTAACGAGAGTATCGATAGGGA 307

Qy         297 CAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGG 356
               |   | || | || | |   ||  |   | | | ||| | |  ||| ||       |
Db         308 AATAAAGGCTATTAAAACATTAAAAGAACAAAGAATTAGGGGAATTATAATTTGTCCAAC 367

Qy         357 TTCTGGAAATAAGGCCAATATTACCCAGTAC 387
               ||||  | ||  |   | || | |||||
Db         368 CTCTGTTGAAAATGATTTAAATAGCGAGTAC 398
```

```
RESULT 2
US-10-765-790-79/c
; Sequence 79, Application US/10765790
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79
;    LENGTH: 401433
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-765-790-79

  Query Match                 7.4%;  Score 33.8;  DB 6;  Length 401433;
  Best Local Similarity   50.3%;  Pred. No. 2.6;
  Matches   83;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy          5 AAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGA 64
              |||| | || | || || ||| |  |  |  |||||  |  ||||||  |||  |
Db     270713 AGAAATATTTTTTCTTTTTTGTAGTATGTTTACTGAAAACTACTGGGAACAAAACATGGT 270654

Qy         65 CAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTT 124
              ||| | |      ||  || | | ||| ||| |||   | ||| | || || | | |
Db     270653 CAACTGAACTAATGCTTCCTCCTGAAAATATGTTTTTAGAAAAAGTATAAACCCACAAAT 270594

Qy        125 CATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATA 169
              |   | ||||   |||      |  | | | | ||  | |||||
Db     270593 ACTAAGAACAGGAAACCAGATGAGAGTATGTGGAAACATAGAATA 270549


RESULT 3
US-10-779-543-9321
; Sequence 9321, Application US/10779543
; GENERAL INFORMATION:
;  APPLICANT: Williams et al
;  TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
;  TITLE OF INVENTION:  IN CANCEROUS CELLS AND THEIR METHODS OF USE II
;  FILE REFERENCE: 2300-21302
;  CURRENT APPLICATION NUMBER: US/10/779,543
;  CURRENT FILING DATE:  2004-02-12
;  PRIOR APPLICATION NUMBER: 10/076,555
;  PRIOR FILING DATE: 2002-02-15
;  PRIOR APPLICATION NUMBER: 09/217,471
;  PRIOR FILING DATE: 1998-12-21
;  PRIOR APPLICATION NUMBER: 60/068,755
;  PRIOR FILING DATE: 1997-12-23
;  PRIOR APPLICATION NUMBER: 60/080,664
;  PRIOR FILING DATE: 1998-04-03
;  PRIOR APPLICATION NUMBER: 60/105,234
;  PRIOR FILING DATE: 1998-10-21
;  PRIOR APPLICATION NUMBER: 09/297,648
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: PCT/US99/01619
;  PRIOR FILING DATE: 1999-01-28
;  PRIOR APPLICATION NUMBER: 60/072,910
;  PRIOR FILING DATE: 1998-01-28
;  PRIOR APPLICATION NUMBER: 60/075,954
;  PRIOR FILING DATE: 1998-02-24
;  PRIOR APPLICATION NUMBER: 60/080,114
;  PRIOR FILING DATE: 1998-03-31
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 23767
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9321
;    LENGTH: 418
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-779-543-9321

  Query Match                 7.2%;  Score 32.8;  DB 6;  Length 418;
  Best Local Similarity   53.0%;  Pred. No. 0.74;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy         87 AGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTAT 146
              ||||| | |  || |  ||||  | |  | ||| || | | ||| | | | | ||
Db        208 AGAATTTCAGACTGGGCGCCATGACCAAGGGCAATCTGCAGTCACTCAAGAGTACTTGGA 267

Qy        147 TGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGTC 206
              | |  | |  ||| | | | |  |||| | ||||||||||  | ||  | |
Db        268 AGATTGATTTAGCAGTGGTGCTTGGAACAGACTGGGCCAGGAAGGGCTATAAGCAGGGAG 327

Qy        207 CGTTATTTCACA 218
              || |||| | |
Db        328 AGTCATTTAAAA 339


RESULT 4
US-10-767-471-8836/c
; Sequence 8836, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8836
;    LENGTH: 201
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-767-471-8836

  Query Match                 7.1%;  Score 32.4;  DB 6;  Length 201;
  Best Local Similarity   55.6%;  Pred. No. 0.8;
  Matches   60;  Conservative    1;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              |||| ||  |  || |  ||||  | |||  ||  |||:|||   |||| |||||
Db        142 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 83

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |  ||  |  |  | | || |||| |||||   |  |||
Db         82 AGGATTATCATGTACATGATGAGGCAGAGCAGGAACAGAAAAAGCTGG 35


RESULT 5
US-10-767-471-34493/c
; Sequence 34493, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34493
;    LENGTH: 201
;    TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-10-767-471-34493

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 201;
  Best Local Similarity   55.6%;  Pred. No. 0.8;
  Matches   60;  Conservative    1;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | || | ||  |  || |  |||||  | |||  ||  |||:|||    |||| |||||
Db        142 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 83

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |   ||   |   |  |  | || |||| |||||    |  |||
Db         82 AGGATTATCATGTACATGATGAGGCAGAGCAGGAACAGAAAAAGCTGG 35


RESULT 6
US-10-767-471-501/c
; Sequence 501, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 501
;   LENGTH: 1227
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-501

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 1227;
  Best Local Similarity   55.6%;  Pred. No. 1.3;
  Matches   60;  Conservative    1;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | || | ||  |  || |  |||||  | |||  ||  |||:|||    |||| |||||
Db        476 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 417

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |   ||.  |   |  |  | || |||| |||||    |  |||
Db        416 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 369


RESULT 7
US-10-767-471-10718/c
; Sequence 10718, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10718
;   LENGTH: 76573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(76573)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10718

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 76573;
  Best Local Similarity   55.6%;  Pred. No. 4.4;
  Matches   60;  Conservative    1;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | || | ||  |  || |  |||||  | |||  ||  |||:|||    |||| |||||
Db      69821 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 69762

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |   ||   |   |  |  | || |||| |||||    |  |||
Db      69761 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 69714


RESULT 8
US-10-767-795-2887
; Sequence 2887, Application US/10767795
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Zhou, Yihua
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53534)B
;  CURRENT APPLICATION NUMBER: US/10/767,795
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 2887
;   LENGTH: 1310
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C71322_1
US-10-767-795-2887

  Query Match             7.0%;  Score 32;  DB 6;  Length 1310;
  Best Local Similarity   55.4%;  Pred. No. 1.8;
  Matches   62;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              |||   |  ||||||    |||| | ||  || | ||  ||    |||  || |   |
Db        626 TGATGCATCCATTGTTCCCAATGAGGATGGGAACATTGCCTGAAAATGGCTATGAATGGG 685

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATT 113
               | ||||||   | || ||    ||| |  ||||   |||   ||||||| |
Db        686 TGTCAAATTCCAAGCTTGCCAACTCAAACGATAACAATGCTCAAAATGAAGT 737


RESULT 9
US-10-775-169-342/c
; Sequence 342, Application US/10775169
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Burczynski, Michael
;  APPLICANT:  Twine, Natalie
;  APPLICANT:  Dorner, Andrew
;  APPLICANT:  Trepicchio, William
;  TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
;  FILE REFERENCE: AM101080 (031896-013000)
;  CURRENT APPLICATION NUMBER: US/10/775,169
;  CURRENT FILING DATE:  2004-02-11
;  NUMBER OF SEQ ID NOS: 5278
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 342
;   LENGTH: 5495
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-169-342

  Query Match             7.0%;  Score 31.8;  DB 6;  Length 5495;
  Best Local Similarity   51.0%;  Pred. No. 3.2;
  Matches   75;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              || ||||||   ||||  |      ||   | ||     |   |         || |   | |
Db       4776 TGCAAAACAGGATGTTTCTCCGTTTGAAAAAAAACAAGCAAGGAAAACAGAGTCATATCA 4717

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
```

```
Db        4716 CGTAAGCGCTTGTTCAGGGAAAATATGACCTTAAGTATGCTGTAAATGGATTTGAAAAGG 4657

Qy         122 CTTCATTTAATCAGGCGGCCATTATTG 148
               || | || | | |||||| | ||  ||
Db        4656 CTCCTTTGAGTTAGGCGGGCTTTTGTG 4630


RESULT 10
US-10-767-471-8837/c
; Sequence 8837, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8837
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-8837

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 201;
  Best Local Similarity   55.6%;  Pred. No. 1.9;
  Matches   60;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | || | ||  |  || |  |||||  | |||  ||  ||| |||     |||| |||||
Db        183 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATGAGGAGGCTATTTCCCAGG 124

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |   ||   |   |  |  || |||| |||||     |  |||
Db        123 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 76


RESULT 11
US-10-767-471-34439/c
; Sequence 34439, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34439
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-34439

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 201;
  Best Local Similarity   55.6%;  Pred. No. 1.9;
  Matches   60;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | || | ||  |  || |  |||||  | |||  ||  ||| |||     |||| |||||
Db        183 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATGAGGAGGCTATTTCCCAGG 124

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |   ||   |   |  |  || |||| |||||     |  |||
Db        123 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 76


RESULT 12
US-60-548-091-5725
; Sequence 5725, Application US/60548091
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STROKE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001506
;  CURRENT APPLICATION NUMBER: US/60/548,091
;  CURRENT FILING DATE:  2004-02-27
;  NUMBER OF SEQ ID NOS: 24433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5725
;   LENGTH: 394468
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(394468)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-60-548-091-5725

  Query Match             6.8%;  Score 31;  DB 7;  Length 394468;
  Best Local Similarity   57.9%;  Pred. No. 19;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        233 ATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCA 292
              |||    || ||| ||  |||| | || |      ||||||  |    | ||||||
Db     320889 ATCTCTTGATAGTTGAGGAGGCTGTGAGTATACGCTTTGCTCACCAGGTGCAAACAGCTG 320948

Qy        293 ATGCCAACGATGCCAGTATATCGCAAAGCGCTTAC 327
              || |||  |  ||||||  || | ||| | ||
Db     320949 GTGTCAATAACCCCAGTAAGTCCCCAAGTGTTTCC 320983


RESULT 13
US-10-257-047-43
; Sequence 43, Application US/10257047
; GENERAL INFORMATION:
;  APPLICANT: STEWARD, MICHAEL
;  APPLICANT:  COX, VIVIENNE FRANCES
;  TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
;  FILE REFERENCE: 37945-0041
;  CURRENT APPLICATION NUMBER: US/10/257,047
;  CURRENT FILING DATE:  2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/GB01/01599
;  PRIOR FILING DATE: 2001-04-09
;  PRIOR APPLICATION NUMBER: GB 0008582.9
;  PRIOR FILING DATE: 2000-04-08
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
;   LENGTH: 309
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-10-257-047-43

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 309;
  Best Local Similarity   53.3%;  Pred. No. 2.8;
  Matches   65;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              ||| || | ||    ||  | | | |   ||||   ||| |||| |   ||   || |
Db        132 TGAGAACCCAAACCCTACCTGTAACGAGAACAACGGTGGATGCGACGCTGACGCTAAGTG 191

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
              | | ||  |  ||||| ||  | |||   | ||    |  |||| | ||||| ||
Db        192 CACCGAAGAAGACTCTGGTTCTAACGGAAAGAAGATTACTTGCGAATGTACTAAGCCAGA 251

Qy        122 CT 123
              ||
Db        252 CT 253


RESULT 14
US-10-257-047-46
```

```
; Sequence 46, Application US/10257047
; GENERAL INFORMATION:
;  APPLICANT: STEWARD, MICHAEL
;  APPLICANT:  COX, VIVIENNE FRANCES
;  TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
;  FILE REFERENCE: 37945-0041
;  CURRENT APPLICATION NUMBER: US/10/257,047
;  CURRENT FILING DATE:  2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/GB01/01599
;  PRIOR FILING DATE: 2001-04-09
;  PRIOR APPLICATION NUMBER: GB 0008582.9
;  PRIOR FILING DATE: 2000-04-08
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
;   LENGTH: 309
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-10-257-047-46

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 309;
  Best Local Similarity   53.3%;  Pred. No. 2.8;
  Matches   65;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              ||| || | ||    ||  | | | |   ||||   ||| |||| |   ||    ||  |
Db        132 TGAGAACCCAAACCCTACCTGTAACGAGAACAACGGTGGATGCGACGCTGACGCTAAGTG 191

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
              |  | ||  |  ||||| ||   | |||   |  ||    |  |||| | ||||| ||
Db        192 CACCGAAGAAGACTCTGGTTCTAACGGAAAGAAGATTACTTGCGAATGTACTAAGCCAGA 251

Qy        122 CT 123
              ||
Db        252 CT 253


RESULT 15
US-10-257-047-44
; Sequence 44, Application US/10257047
; GENERAL INFORMATION:
;  APPLICANT: STEWARD, MICHAEL
;  APPLICANT:  COX, VIVIENNE FRANCES
;  TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
;  FILE REFERENCE: 37945-0041
;  CURRENT APPLICATION NUMBER: US/10/257,047
;  CURRENT FILING DATE:  2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/GB01/01599
;  PRIOR FILING DATE: 2001-04-09
;  PRIOR APPLICATION NUMBER: GB 0008582.9
;  PRIOR FILING DATE: 2000-04-08
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Vector
;   OTHER INFORMATION: pVK104-01
US-10-257-047-44

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 3147;
  Best Local Similarity   53.3%;  Pred. No. 5.5;
  Matches   65;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              ||| || | ||    ||  | | | |   ||||   ||| |||| |   ||    ||  |
Db        198 TGAGAACCCAAACCCTACCTGTAACGAGAACAACGGTGGATGCGACGCTGACGCTAAGTG 257

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
              |  | ||  |  ||||| ||   | |||   |  ||    |  |||| | ||||| ||
Db        258 CACCGAAGAAGACTCTGGTTCTAACGGAAAGAAGATTACTTGCGAATGTACTAAGCCAGA 317

Qy        122 CT 123
              ||
Db        318 CT 319


Search completed: March 19, 2004, 03:09:57
Job time : 56.1445 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		132.2	29.0	220	12	BI941542	BI941542 dg26g06.y
2	c	53.6	11.8	207	9	AI368352	AI368352 SWOvMfCAR
3		39.4	8.6	1101	29	CNS00E93	AL068795 Drosophil
4	c	39	8.6	680	12	BJ618688	BJ618688 BJ618688
5		38.8	8.5	775	9	AI664697	AI664697 TENG0665
6	c	38	8.3	288	28	AZ578608	AZ578608 25d10 Sho
7		37.8	8.3	675	12	BP007561	BP007561 BP007561
8	c	37.8	8.3	715	9	AV896333	AV896333 AV896333
9		37.6	8.2	601	13	BU213938	BU213938 603108484
10	c	37.4	8.2	1214	28	CC310289	CC310289 TAM32-32B
11		36.8	8.1	708	13	BU387712	BU387712 603857374
12		36.8	8.1	924	28	AZ529672	AZ529672 ENTBM10TR
13	c	36.4	8.0	1550	28	CC211905	CC211905 CH261-186
14	c	36	7.9	499	28	BH528896	BH528896 BOGPO61TR
15		36	7.9	579	12	BG226083	BG226083 kq08h12.y
16	c	36	7.9	795	28	BZ229556	BZ229556 CH230-353
17	c	36	7.9	797	28	BH505143	BH505143 BOGBM31TF
18		36	7.9	975	29	CNS03TIU	AL259887 Tetraodon
19		35.8	7.9	883	28	BH159159	BH159159 ENTRB03TR
20		35.6	7.8	412	10	AW140314	AW140314 SWAMCAC45
21	c	35.6	7.8	760	28	BZ756658	BZ756658 PUFBN68TD
22		35.2	7.7	659	12	BJ481788	BJ481788 BJ481788
23	c	35.2	7.7	791	14	CB310770	CB310770 AGENCOURT
24		34.8	7.6	719	12	BG548059	BG548059 602575980
25		34.8	7.6	1201	13	BX406892	BX406892 BX406892
26		34.6	7.6	606	10	BE581316	BE581316 kq49c03.y
27	c	34.6	7.6	698	28	BH943841	BH943841 odf17b10.
28		34.6	7.6	700	13	BW245005	BW245005 BW245005
29		34.6	7.6	710	9	AV940831	AV940831 AV940831
30		34.4	7.5	435	14	CA856583	CA856583 PfESToac1
31		34.4	7.5	471	28	AQ055856	AQ055856 CIT-HSP-2
32		34.4	7.5	576	12	BM273661	BM273661 PfESToaa5
33		34.4	7.5	577	13	BQ451444	BQ451444 PfESToaa9
34		34.4	7.5	581	14	CB389390	CB389390 OSTF116E9
35		34.4	7.5	633	12	BJ104432	BJ104432 BJ104432
36	c	34.4	7.5	735	28	BZ110007	BZ110007 CH230-164
37		34.4	7.5	1094	29	CNS00BR5	AL057531 Drosophil
38	c	34.2	7.5	666	13	BU582719	BU582719 rd26b04.y
39	c	34.2	7.5	689	13	BQ625251	BQ625251 rd26b04.y
40	c	34.2	7.5	784	28	BZ280200	BZ280200 CH230-297
41		34.2	7.5	854	12	BM012028	BM012028 603636451
42		34.2	7.5	965	9	AL549572	AL549572 AL549572
43		34.2	7.5	1084	28	CC299970	CC299970 CH261-89H
44		34.2	7.5	1224	14	CF238215	CF238215 AGENCOURT
45		34	7.5	402	9	AW043435	AW043435 SWAMCAC37

ALIGNMENTS

RESULT 1
BI941542
LOCUS BI941542 220 bp mRNA linear EST 19-FEB-2003
DEFINITION dg26g06.y1 Xenopus laevis gastrula non normalized Xenopus laevis cDNA clone IMAGE:3749963 5' similar to SW:CSGB_ECOLI P39828 MINOR CURLIN SUBUNIT PRECURSOR. ;, mRNA sequence.
ACCESSION BI941542
VERSION BI941542.1 GI:16256014
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 220)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Source lab clone id - 3749963
Trace considered overall poor quality This clone is available royalty-free through LLNL; contact the IMAGE Consortium (image@image.llnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 1.

FEATURES
source Location/Qualifiers
1. .220
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:3749963"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F'"
/clone_lib="Xenopus laevis gastrula non normalized"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2 ug of poly A+ RNA (equal parts from stage 10.5 and stage 11.5 gastrulae). EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. The library was mass excised and used to infect Top10F'. Clones were picked into freezing medium (per liter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM K2HPO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4 7 H2O, 6.8 mM (NH4)2SO4, 4 % w/v glycerol) and grown for 24 hours. Original library construction by Bruce Blumberg (Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

Query Match 29.0%; Score 132.2; DB 12; Length 220;
Best Local Similarity 75.6%; Pred. No. 5.2e-30;
Matches 164; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy        80 CTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGG 139
             ||  |||| |||||||  || ||||  | |||||||  ||  ||||||||||||||||||| |
Db         3 CTAATTCATAATATAACTTCGCGGCACATGAATCGAGTAAGTCTTCATTTAATCAGGCAG 62

Qy       140 CCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAAC 199
             ||||  ||||||||||   ||  ||  |||||||||        |||  |||| ||| |||||||
Db        63 CCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACTGCAGGGAGGCTCAAAAC 122

Qy       200 TATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGA 259
             | ||| | |||   | | ||||||||  | |  || |||||  ||   | |||||| ||||
Db       123 TTTTGGCGGTTGCTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAG 182

Qy       260 ATTATAACTTTGCGTATATTGAGCAAACGGGCAATGC 296
             ||||||||  ||||  ||||||||  ||   |||||| | |
Db       183 ATTATAACCTTGCATATATTGATCAGGCGGGCAGTCC 219
```

RESULT 2
AI368352/c
LOCUS AI368352 207 bp mRNA linear EST 11-JAN-1999
DEFINITION SWOvMfCAR04G09SK Onchocerca volvulus microfilaria cDNA (SAW98MLW-OvMf) Onchocerca volvulus cDNA clone SWOvMfCAR04G09 5', mRNA sequence.
ACCESSION AI368352
VERSION AI368352.1 GI:4147105
KEYWORDS EST.
SOURCE Onchocerca volvulus
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 207)
AUTHORS Williams,S.A.
TITLE Genes expressed in microfilaria of Onchocerca volvulus
JOURNAL Unpublished (1999)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.

FEATURES
source Location/Qualifiers
1. .207
/organism="Onchocerca volvulus"
/mol_type="mRNA"
/db_xref="taxon:6282"
/clone="SWOvMfCAR04G09"
/dev_stage="microfilaria"
/lab_host="XL1-Blue MRF'"
/clone_lib="Onchocerca volvulus microfilaria cDNA (SAW98MLW-OvMf)"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.8 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email:genome@smith.edu."

ORIGIN

Query Match 11.8%; Score 53.6; DB 9; Length 207;
Best Local Similarity 91.8%; Pred. No. 1.4e-05;
Matches 56; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy         1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
             |||||||  | ||||||  |||||||||||| ||||||||||||||||||||||||||| ||
Db        61 ATGAAAACCNAATTGTCATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 2

Qy        61 G 61
             |
Db         1 G 1
```

RESULT 3
CNS00E93
LOCUS CNS00E93 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR28G06 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL068795
VERSION AL068795.1 GI:4949039
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR28G06"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

Query Match 8.6%; Score 39.4; DB 29; Length 1101;
Best Local Similarity 33.2%; Pred. No. 0.61;
Matches 92; Conservative 52; Mismatches 133; Indels 0; Gaps 0;

Qy 2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
Db 825 KGAAAAAAWCCTTGATAAAYGTGAWAATWDAAAAASKGTTTCTSTCTTTTTTWDTTGGGS 884

Qy 62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
Db 885 TGCMTAAATATYKRGGGGAAGRGKTAAARDTSGDAKKTRTTATTBSATSMBTWAATTARB 944

Qy 122 CTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTAC 181
Db 945 KSTAAWTTKGSCCKAWCCCCTTTTTTTBTTTTTTTTKGGVGGGAAWWTTTKKTTKGTTTT 1004

Qy 182 GCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGA 241
Db 1005 TKKARGGAGGGGGRAAAAATTTGGAATGWAATDAARAWRAGGRAGRRARAKAARGGRGGA 1064

Qy 242 AAGTCGACCAGGCAGGGAATTATAACTTTGCGTATAT 278
Db 1065 AAWRGGAAAAGATADAAWTKGTWTTTTGGRTGAWWAT 1101

RESULT 4
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.
http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 8.6%; Score 39; DB 12; Length 680;
Best Local Similarity 89.4%; Pred. No. 0.69;
Matches 42; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 410 TAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
Db 680 TAGTGCAGAGACAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 634

RESULT 5
AI664697
LOCUS AI664697 775 bp mRNA linear EST 07-JUL-1999
DEFINITION TENG0665 T. Cruzi epimastigote normalised cDNA Library Trypanosoma cruzi cDNA clone n1059.r 5', mRNA sequence.
ACCESSION AI664697
VERSION AI664697.1 GI:4775685
KEYWORDS EST.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 775)
AUTHORS Gonzalez Rey,E., Remisz,E., Delgado Garcia,A. and Gonzalez,A.
TITLE Characterization of ESTs from Trypanosoma cruzi epimastigotes
JOURNAL Unpublished (1998)
COMMENT Contact: Delgado Alberto
Departamento de Biologia Molecular, Lab 303
Instituto de Parasitologia y Biomedicina
Consejo Superior de Investigaciones Cientificas C/ Ventanilla No 11, E-18001, Granada, Spain
Tel: 34 958 805058
Fax: 34 958 203323
Seq primer: T7
High quality sequence stop: 775.
FEATURES Location/Qualifiers
source 1. .775
/organism="Trypanosoma cruzi"
/mol_type="mRNA"
/strain="Cl - Brenner"
/db_xref="taxon:5693"
/clone="n1059.r"
/cell_type="epimastigote"
/clone_lib="T. Cruzi epimastigote normalised cDNA Library"
/note="Site_1: EcoRI; Site_2: NotI; cDNA library constructed with oligo dt primed epimastigote mRNA and cloned in pt7t318D phagemid with modified polylinker"
ORIGIN

Query Match 8.5%; Score 38.8; DB 9; Length 775;
Best Local Similarity 57.4%; Pred. No. 0.84;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 245 TCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATG 304
Db 97 TCGCAGTGGCAGCGAATTCGAGATTTGCATTCTGGGAGAAACACGGTCATGCCGTCTGTT 156

Qy 305 CCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAA 364
Db 157 CCTGTAGATCCAACACTGTCGTCGGTGCGACCTCAGTCATTACACACGAAGCTGGTGGAA 216

Qy 365 AT 366
Db 217 AT 218

RESULT 6

```
AZ578608/c
LOCUS       AZ578608                 288 bp    DNA     linear   GSS 08-DEC-2000
DEFINITION  25d10 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
            sp. NGR234 genomic clone 25d10, genomic survey sequence.
ACCESSION   AZ578608
VERSION     AZ578608.1  GI:11606060
KEYWORDS    GSS.
SOURCE      Rhizobium sp. NGR234
  ORGANISM  Rhizobium sp. NGR234
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
REFERENCE   1  (bases 1 to 288)
  AUTHORS   Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
  TITLE     Genetic snapshots of the Rhizobium species NGR234 genome
  JOURNAL   Genome Biol. 1 (6), RESEARCH0014 (2000)
  MEDLINE   21114532
    PUBMED  11178268
COMMENT     Contact: Virginie Viprey
            Laboratoire de Biologie Moleculaire des Plantes Superieures
            University of Geneva
            1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
            Tel: +44(0)1603450000
            Fax: +44(0)1603450045
            Email: virginie.viprey@bbsrc.ac.uk
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism="Rhizobium sp. NGR234"
                     /mol_type="genomic DNA"
                     /strain="ANU265"
                     /db_xref="taxon:394"
                     /clone="25d10"
                     /clone_lib="Shot-gun genomic library of Rhizobium strain
                     ANU265"
                     /note="Vector: M13; derivative strain of NGR234 cured of
ORIGIN               pNGR234a"

  Query Match                 8.3%;  Score 38;  DB 28;  Length 288;
  Best Local Similarity   50.0%;  Pred. No. 1.1;
  Matches   95;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        204 GTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTA 263
              ||||||   |  || |||||          |||  |  |  |  |||     || ||  |
Db        238 GTCCGTGTCTGTTCAGGAAGGCAAGGGCAATATCTCCGGAACGGCGCAGTTCGGAAAGAA 179

Qy        264 TAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGC 323
              ||  |||   |  |||||||||||||  |||||  |  |  |  |  ||   ||
Db        178 CCACAAGGCGATCACCCTGCAAAAGGGCAAGTACAACGCTTCGGGCACGTTCCAGTTCGG 119

Qy        324 TTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCA 383
              ||||||  ||  |   |    ||||  ||   ||||||   |||  |    ||
Db        118 TTACGGCAACAAATCGCTGACCCACCAGGACGGAGAAGGAAATCTCTCCATCACAGTGCA 59

Qy        384 GTACGGTACG 393
              |  ||  |  ||
Db         58 GGACTGATCG 49


RESULT 7
BP007561
LOCUS       BP007561                 675 bp    mRNA    linear   EST 15-MAR-2002
DEFINITION  BP007561 Nori Satoh unpublished cDNA library, young adult Ciona
            intestinalis cDNA clone ciad45a18 5', mRNA sequence.
ACCESSION   BP007561
VERSION     BP007561.1  GI:19499038
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .675
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="ciad45a18"
                     /tissue_type="whole animal"
                     /dev_stage="young adult"
                     /clone_lib="Nori Satoh unpublished cDNA library, young
ORIGIN               adult"

  Query Match                 8.3%;  Score 37.8;  DB 12;  Length 675;
  Best Local Similarity   50.3%;  Pred. No. 1.6;
  Matches   93;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        229 AATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACG 288
              ||||  |          ||   ||  |  |||  |  |  |  ||  |  |  |  |||  |||
Db        257 AATACTTTCAATGAAAAAGAAATGTCAGAAACTAAAACCTGTACTGAAACAGAGAAAAAA 316

Qy        289 GGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATC 348
              ||   ||   |  |  ||  |  |||   ||  |  |  |  |  |||  ||  ||
Db        317 TGCTCCACCTCTGCAGGCAATGTATTAGATACAAAAGAAAGCAATTTAGCTTCTGAAACT 376

Qy        349 CAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTT 408
              ||  |  |||||  ||||||  |  ||||  |      |||  |  |  ||  |  |||  |||
Db        377 CAAATAATTTCTGCAAATAATGACAATTTCCAGCAGAATGATAACGACACTACACCAGAT 436

Qy        409 GTAGT 413
              |  |  |
Db        437 GAACT 441


RESULT 8
AV896333/c
LOCUS       AV896333                 715 bp    mRNA    linear   EST 09-NOV-2001
DEFINITION  AV896333 Nori Satoh unpublished cDNA library, young adult Ciona
            intestinalis cDNA clone rciad45a18 3', mRNA sequence.
ACCESSION   AV896333
VERSION     AV896333.1  GI:16885431
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="rciad45a18"
                     /tissue_type="whole animal"
                     /dev_stage="young adult"
```

/clone_lib="Nori Satoh unpublished cDNA library, young adult"

ORIGIN

```
  Query Match             8.3%;  Score 37.8;  DB 9;  Length 715;
  Best Local Similarity   50.3%;  Pred. No. 1.7;
  Matches   93;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        229 AATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACG 288
              |||| |       ||   ||   | |||  | | | | || | |  | |  ||| |||
Db        620 AATACTTTCAATGAAAAAGAAATGTCAGAAACTAAAACCTGTACTGAAACAGAGAAAAAA 561

Qy        289 GGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATC 348
               ||    ||   |  | || | |||   | |       | | |||   ||  ||   |
Db        560 TGCTCCACCTCTGCAGGCAATGTATTAGATACAAAAGAAAGCAATTTAGCTTCTGAAACT 501

Qy        349 CAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTT 408
              || | |  ||||| |||||| | |||| |     ||| | | ||   | |  ||| ||| |
Db        500 CAAATAATTTCTGCAAATAATGACAATTTCCAGCAGAATGATAACGACACTACACCAGAT 441

Qy        409 GTAGT 413
              | | |
Db        440 GAACT 436
```

RESULT 9
BU213938
LOCUS BU213938 601 bp mRNA linear EST 25-NOV-2002
DEFINITION 603108484F1 CSEQCHN04 Gallus gallus cDNA clone ChEST51g22 5', mRNA sequence.
ACCESSION BU213938
VERSION BU213938.1 GI:25389321
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES Location/Qualifiers
source 1. .601
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST51g22"
/tissue_type="whole embryo"
/dev_stage="20-21"
/lab_host="DH10B"
/clone_lib="CSEQCHN04"
/note="Organ: whole embryo; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: Not1; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

```
  Query Match             8.2%;  Score 37.6;  DB 13;  Length 601;
  Best Local Similarity   50.6%;  Pred. No. 1.8;
  Matches   91;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy         87 AGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTAT 146
              || | | ||| |    |||||||  |  |  | || ||  |  | | |  | | |  | ||
Db        172 AGCAAAAAATCTACGCGTAAATGCAGGAGACCAGCCTGGAGCTGACCTAGGGCCTCTAAT 231

Qy        147 TGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGTC 206
               |||      | || ||| |     |  |     |     ||| |||||  | ||  |     |
Db        232 CAGTCCCCAAGCCAAGGAAAGGGTTTGCCACCTGATTCAGAAAGGAGTAGAAGAAGGTGC 291

Qy        207 CGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAA 266
              |  | ||   |  || ||| | |||| |   | ||| | | |  ||   || |||| | |
Db        292 CAGTCTTCTTCTGGATGGACGTAATATTAAAGTGAAGGGCTATGAGAACGGCAATTCTGA 351
```

RESULT 10
CC310289/c
LOCUS CC310289 1214 bp DNA linear GSS 14-MAY-2003
DEFINITION TAM32-32B18_EC1.1 TAM32 Gallus gallus genomic clone TAM32-32B18, genomic survey sequence.
ACCESSION CC310289
VERSION CC310289.1 GI:30703437
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1214)
AUTHORS Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 152000 Std Error: 0.00
Seq primer: EC1 TACGACTCACTATAGGGCG
Class: BAC ends
High quality sequence start: 20
High quality sequence stop: 781.
FEATURES Location/Qualifiers
source 1. .1214
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="TAM32-32B18"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="TAM32"
/note="Vector: pECBAC1; Site_1: EcoRI; Site_2: EcoRI; TAM32 Female Chicken library - for library and clone ordering information: http://www.hbz.tamu.edu"

ORIGIN

```
  Query Match             8.2%;  Score 37.4;  DB 28;  Length 1214;
  Best Local Similarity   47.3%;  Pred. No. 2.6;
  Matches  113;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy         71 ATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTA 130
              ||  | ||| | ||||| | || | |||    ||  | | |  | | |||| ||||   |
Db       1027 ATATTTTGGATGGTTCAAAGTAAATTTTGAGGGGCATTTACAGATGGAAGTTTTCAAAAA 968
```

```
Qy        131 ATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAG 190
Db        967 ATCCAGGGAACTTTATTAGTGGAGTAGAGAACCATCCTGGTTAATTCCTGGTCCTTGAGT 908

Qy        191 GATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACC 250
Db        907 TAGCAAAATGGGATACCTTAAGAGAACAAAAGAAAACCTGGCCATTTGCGGTCTCCCAAA 848

Qy        251 AGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGT 309
Db        847 TACCAGGAAACAAACTATTTATCTATCCTGCTCAAACAGTTCAAGGAAGCTAATCTAAT 789
```

RESULT 11
BU387712

LOCUS	BU387712 708 bp mRNA linear EST 28-NOV-2002
DEFINITION	603857374F1 CSEQCHN75 Gallus gallus cDNA clone ChEST863c20 5', mRNA sequence.
ACCESSION	BU387712
VERSION	BU387712.1 GI:25895713
KEYWORDS	EST.
SOURCE	Gallus gallus (chicken)
ORGANISM	Gallus gallus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE	1 (bases 1 to 708)
AUTHORS	Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE	A Comprehensive Collection of Chicken cDNAs
JOURNAL	Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE	22335534
PUBMED	12445392
COMMENT	Contact: Simon Hubbard Department of Biomolecular Sciences University of Manchester Institute of Science and Technology (UMIST) PO Box 88, Manchester, M60 1QD, UK Tel: 01612008930 Fax: 01612360409 Email: Simon.Hubbard@umist.ac.uk.

FEATURES Location/Qualifiers

source 1. .708
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST863c20"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHN75"
/note="Organ: trunks; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: Not1; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

```
Query Match             8.1%;  Score 36.8;  DB 13;  Length 708;
Best Local Similarity   50.6%;  Pred. No. 3.4;
Matches   89;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         89 AATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTATTG 148
Db        17 AAGAGAATCTACGCGTAAATGCAGGAGACCAGCCTGGAGCTGACCTAGGGCCTCTAATCA 76

Qy        149 GTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGTCCG 208
Db         77 GTCCCCAAGCCAAGGAAAGGGTTTGCCACCTGATTCAGAAAGGAGTAGAAGAAGGTGCCA 136

Qy        209 TTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
Db        137 GTCTTCTTCTGGATGGACGTAATATTAAAGTGAAGGGCTATGAGAACGGCAATTTT 192
```

RESULT 12
AZ529672

LOCUS	AZ529672 924 bp DNA linear GSS 03-NOV-2000
DEFINITION	ENTBM10TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
ACCESSION	AZ529672
VERSION	AZ529672.1 GI:11082756
KEYWORDS	GSS.
SOURCE	Entamoeba histolytica
ORGANISM	Entamoeba histolytica Eukaryota; Entamoebidae; Entamoeba.
REFERENCE	1 (bases 1 to 924)
AUTHORS	Loftus,B., Van Aken,S. and Fraser,C.
TITLE	Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: bjloftus@tigr.org Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library Seq primer: M13-Reverse Class: shotgun High quality sequence start: 33 High quality sequence stop: 745.

FEATURES Location/Qualifiers

source 1. .924
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

ORIGIN

```
Query Match             8.1%;  Score 36.8;  DB 28;  Length 924;
Best Local Similarity   50.6%;  Pred. No. 3.7;
Matches   89;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         73 GATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAAT 132
Db        740 GATGTTACACTTAGAAGATTTATATAAGGGGTAAAACTATTAGAGGGTATTACATTAGAA 799

Qy        133 CAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGA 192
```

```
Db        800 AAAGACACATTATTTGTCATCATTGTCATGGGTACTGGGAAGGAGATACACCACGTATGA 859

Qy        193 TCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGA 248
              |  |||  | |||||  ||| |||   ||||  || | | |    |||| |   ||
Db        860 TGTAAAGAAATGTCCAGTATGTCATTGGAAGATGGGAGTTAAAACGCGAGACAAGA 915


RESULT 13
CC211905/c
LOCUS       CC211905               1550 bp    DNA     linear   GSS 12-MAY-2003
DEFINITION  CH261-186O6_Sp6.1 CH261 Gallus gallus genomic clone CH261-186O6,
            genomic survey sequence.
ACCESSION   CC211905
VERSION     CC211905.1  GI:30530573
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1550)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 339
            High quality sequence stop: 1106.
FEATURES             Location/Qualifiers
     source          1. .1550
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-186O6"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 28;  Length 1550;
  Best Local Similarity   59.8%;  Pred. No. 5.9;
  Matches   61;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         97 TTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTC 156
              |||| |   | ||||||| |||  || ||| | ||||  ||  ||    |||| ||   ||
Db       1185 TTTGGGTGGACTGAATTGAGCCTGTGTTCTTGTAATGTGGAAGCACAAATTGTTCCTCTC 1126

Qy        157 GGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAA 198
              || | |  ||||   |   ||| |    |||||||  ||||
Db       1125 AGCTCAGGAAATATCTCTGTAGTTCAGGGGGAAGGAAAAAAA 1084


RESULT 14
BH528896/c
LOCUS       BH528896               499 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGPO61TR BOGP Brassica oleracea genomic clone BOGPO61, genomic
            survey sequence.
ACCESSION   BH528896
VERSION     BH528896.1  GI:17744694
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGPO61TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGPO61"
                     /clone_lib="BOGP"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             7.9%;  Score 36;  DB 28;  Length 499;
  Best Local Similarity   46.4%;  Pred. No. 5.4;
  Matches  117;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAA 242
              || | || || ||||   |||  | || |||  | | | ||  || || |        |
Db        470 CCGGAAACGAGCAAATGCATTCACGGTGATTGAAGATGCAGAGGGCAACATGGTCTATCA 411

Qy        243 AGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGA 302
              ||  || |||  |||||      | | ||| |||| || |  |       |||  | |  ||
Db        410 AGAAGATCAGATAGGGAGAGTGATCATTGAGTATTTTCATGAGCTATTCAAGACTATTGA 351

Qy        303 TGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGG 362
              | ||  | |  | ||| ||     ||  |    || | ||  |     ||  |
Db        350 GGGCAACAGAGAGAAAACCGTAATGCGTGCTCTATCACCAATGGTGAGCGCAGAAACGAA 291

Qy        363 AAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACA 422
              | |||   | ||    ||   | | | |  |||  | ||  |   || |    |  ||
Db        290 CGAGAAGCTAATTACAGTCCCAAAAGCTGCAGAGATCAAAGAGGCCCTATTCTCTATTCA 231

Qy        423 GTCGCATATGGC 434
              |  | |  ||
Db        230 TGCCGACAAAGC 219


RESULT 15
BG226083
LOCUS       BG226083               579 bp    mRNA    linear   EST 09-MAY-2001
DEFINITION  kq08h12.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to
            WP:ZK686.3 CE00457 ;, mRNA sequence.
ACCESSION   BG226083
VERSION     BG226083.1  GI:12713638
KEYWORDS    EST.
SOURCE      Strongyloides stercoralis
  ORGANISM  Strongyloides stercoralis
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Panagrolaimoidea; Strongyloididae; Strongyloides.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
```

```
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Thomas Nutman and colleagues of
            NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
            University Genome Sequencing Center St. Louis.
            High quality sequence stop: 417.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Strongyloides stercoralis"
                     /mol_type="mRNA"
                     /strain="Rhabditiform larvae obtained from gerbils"
                     /db_xref="taxon:6248"
                     /lab_host="XL-1 Blue MRF' (Stratagene)"
                     /clone_lib="TBN95TM-SSR"
                     /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
                     rhabditiform larvae which had been isolated from gerbils
                     experimentally infected with larvae originally isolated
                     from experimentally infected dogs. cDNA was constructed
                     and, using adaptors, was cloned unidirectionally into the
                     vector from the EcoRI site to the XhoI site. The library
                     has an unamplified titer of 1 x 10E5 pfu/ml and an
                     amplified, undiluted titer of 9 x 10E11 pfu/ml. The
                     average insert size of the unamplifed library is 675 bp
                     (range, 100-1700)."
ORIGIN

  Query Match             7.9%;  Score 36;  DB 12;  Length 579;
  Best Local Similarity   49.0%;  Pred. No. 5.6;
  Matches   96;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
              || ||| | |  ||| | ||| |           | ||| || | ||||    | | | |
Db        271 AAAAAACTTTATTTTGTTATGGTTGATTATGCAGATGCTCCAGAGATTTTTGCTTCAATA 330

Qy         67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
              |||  | ||   ||||   ||       ||||| | |  || | ||  || || |||
Db        331 AATCTTAATACAGCTCCCGCATTTATACATTTTCCAGCTAAAGGATCTAGGAAATCTGAT 390

Qy        127 TTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAG 186
               ||    ||  |      ||| | | | |   | |||  |   || ||| |     |
Db        391 GATACAATGGATGTTCAAATTAATGGATTTGAAGCTGATGTTTTGGCAAGATTTGTTAAA 450

Qy        187 GAAGGATCAAAACTAT 202
              ||| || |  || |||
Db        451 GAAAGAACTGAAATAT 466


Search completed: March 18, 2004, 20:11:11
Job time : 2038.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	222.4	48.8	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	180	39.5	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	33.4	7.3	3990	5	US-09-830-230A-629	Sequence 629, App
	4	32.4	7.1	150223	6	US-10-624-149A-1	Sequence 1, Appli
c	5	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
c	6	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
c	7	29.4	6.4	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	8	29.4	6.4	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	9	29.4	6.4	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	10	29.4	6.4	1400	7	US-60-545-213-6407	Sequence 6407, Ap
c	11	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	12	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
c	13	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	14	28.4	6.2	2511	1	PCT-US04-05654-2123	Sequence 2123, Ap
c	15	28.4	6.2	20043	6	US-10-021-698A-869	Sequence 869, App
	16	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	17	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
c	18	28.2	6.2	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	19	28.2	6.2	3416	6	US-10-100-683-3397	Sequence 3397, Ap
c	20	28.2	6.2	8206	6	US-10-021-698A-3529	Sequence 3529, Ap
c	21	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
c	22	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
c	23	27.8	6.1	639	6	US-10-767-701-7904	Sequence 7904, Ap
	24	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
c	25	27.8	6.1	2342	6	US-10-451-467A-277	Sequence 277, App
	26	27.8	6.1	2895	6	US-10-767-701-14458	Sequence 14458, A
	27	27.8	6.1	18802	6	US-10-767-471-10574	Sequence 10574, A
c	28	27.8	6.1	32372	6	US-10-767-471-10611	Sequence 10611, A
c	29	27.8	6.1	1126118	6	US-10-767-471-10615	Sequence 10615, A
c	30	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
	31	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
	32	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
	33	27.6	6.1	1498	6	US-10-100-683-5108	Sequence 5108, Ap
	34	27.6	6.1	1523	6	US-10-100-683-750	Sequence 750, App
	35	27.6	6.1	1662	6	US-10-100-683-5106	Sequence 5106, Ap
c	36	27.6	6.1	2571	5	US-09-999-183A-4	Sequence 4, Appli
	37	27.6	6.1	3035	6	US-10-786-892-279	Sequence 279, App
c	38	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	39	27.4	6.0	201	6	US-10-767-471-11899	Sequence 11899, A
	40	27.4	6.0	868	6	US-10-767-701-167	Sequence 167, App
c	41	27.4	6.0	318752	6	US-10-767-471-10553	Sequence 10553, A
	42	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
	43	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli
c	44	27	5.9	511	6	US-10-767-701-18763	Sequence 18763, A
c	45	27	5.9	987	6	US-10-100-683-3090	Sequence 3090, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 48.8%; Score 222.4; DB 5; Length 675;
Best Local Similarity 68.0%; Pred. No. 9e-67;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||  |  |||| |  | | | |  ||  ||  ||  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||  | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262
Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442
Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 39.5%; Score 180; DB 5; Length 396;
Best Local Similarity 70.6%; Pred. No. 2.8e-52;
Matches 240; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```
Qy        117 CGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCT 176
Db         57 CGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCT 176
Qy        177 GCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGC 236
Db        117 GCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGC 176
Qy        237 CGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAA 296
Db        177 AGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAA 236
Qy        297 TAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGG 356
```

```
Db        237 CAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGG 296
Qy        357 CGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCA 416
Db        297 TGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCA 356
Qy        417 GGTTGGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        357 GGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 629
; LENGTH: 3990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1139)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1148)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1210)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1244)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1247)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1250)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1251)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

Query Match 7.3%; Score 33.4; DB 5; Length 3990;
Best Local Similarity 47.4%; Pred. No. 0.24;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAG 216
              | ||  |||||| ||||     |    ||||      ||  ||| |  | | |  | | |
Db       3433 GTTACGGCTGCGGCTGCTGGTGAGCAGGATGGAGAGAAGCCTGCAGAGGCTACAAATCCG 3492

Qy        217 AGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |  || ||  ||    | |   |   | ||||||||||  | |||    |||  ||
Db       3493 ATTGCTGCTGCTATTGGGAAGGGTAATGAGGATGGTGCGGATTTTGGTAAGGATGAGATG 3552

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGAT 336
              |   || ||| |   |    |  ||| |          | |||  |||||  |      |
Db       3553 AAGAAGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAG 3612

Qy        337 ATTACTGTCGGCCAATACGGCGGTAATAACG 367
               || ||||       | |  ||| | || |
Db       3613 TTTGCTGTGAAGAGTAATGATGGTGAGAAAG 3643
```

RESULT 4
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
; APPLICANT: Neubauer, Antonie
; APPLICANT: Ziegler, Christina
; TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
; FILE REFERENCE: 1/1372
; CURRENT APPLICATION NUMBER: US/10/624,149A
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 60/403,282
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: DE 10233064
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: DE 10317008
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
; LENGTH: 150223
; TYPE: DNA
; ORGANISM: Equine herpesvirus 1
; PUBLICATION INFORMATION:
; AUTHORS: Telford,E.A.
; AUTHORS: Watson, M.S.
; AUTHORS: McBride, K.
; AUTHORS: Davison, A.J.
; TITLE: The DNA sequence of equine herpesvirus-1
; JOURNAL: Virology
; VOLUME: 189
; ISSUE: 1
; PAGES: 304-316
; DATE: JUL-1992
; DATABASE ACCESSION NUMBER: NC 001491, NCBI
; DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 2.9;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        173 CTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACG 232
              || ||  ||  |  |·|| ||  | ||| ||| |   |  ||| |||   | ||  |||
Db      95542 CTGTGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACG 95601

Qy        233 GCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCA 292
              |       | |||| ||  ||||||    | || |  | | | ||| |    |  |||
Db      95602 TCAAGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCA 95661

Qy        293 GAAATAATGCCACCATCGACCAGTGG 318
              || |   |   || ||| | || |
Db      95662 GAGAGCTTATTCACAACGAGCTGTTG 95687
```

RESULT 5
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.2;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
              |||   | | | | || ||   | | || |  || | |||    || |||| ||||||||
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
              |  | |   | |  |||  |  |   |   |||  |  ||||||||
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583
```

RESULT 6
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10736
; LENGTH: 24900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(24900)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

Query Match 6.5%; Score 29.6; DB 6; Length 24900;
Best Local Similarity 52.4%; Pred. No. 11;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy        207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||  | ||  ||  | ||  | ||  |       |    | ||  ||  |  | |||||  |
Db      10048 CAGCATCCTGACTGCTTTTAGTGTCTCAAAAGTCCTGGGTCAAGTGGTGCATAATTGGTG 9989

Qy        267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAA 326
              |  |||||   |  | | |||||    | ||||| ||| | | | |      ||  |    |||
Db       9988 TCCTGAACCCAAACTGTATGGTACTAAAAATAGTGCAATCCTGGTTTGTTGTATATTTAA 9929

Qy        327 AAAC 330
              || |
Db       9928 AATC 9925
```

RESULT 7

US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2134
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-2134

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db 437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy 272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db 377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy 332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
Db 317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy 392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
Db 257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199

RESULT 8
US-60-545-213-2135/c
; Sequence 2135, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2135
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-2135

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db 437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy 272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db 377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy 332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
Db 317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy 392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
Db 257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199

RESULT 9
US-60-545-213-6406/c
; Sequence 6406, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6406
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-6406

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db 437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

Qy 272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
Db 377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy 332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
Db 317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy 392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
Db 257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199

RESULT 10
US-60-545-213-6407/c
; Sequence 6407, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6407
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-6407

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTG 271
Db 437 CCCTGGGCGATAGCGTTGCAGTAGACGGTGGACCGCAGGTTGGGGTGGATCGGGTTATTA 378

```
Qy        272 AACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACT 331
              |  ||   ||     |      | || ||||    | | |       |        ||||
Db        377 TTGGGGTTCTCCATCCACTGCTTGAAAAGGCCAGAGACCATCTCCTCACACTCTGGAACT 318

Qy        332 CCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCAT 391
              ||| |    | | |||     | |||  |||   ||| |  |||||  |||||      |
Db        317 CCGTTGGAGCAGGCGGTGCTGATGGCATTAACCTCGCTGTACTGGTCCATCAGGTTTTCT 258

Qy        392 CTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAG 450
              |||  |      | | ||  ||    ||  |   | ||  | |    | | ||| |
Db        257 GGGATCTCCCTCCAGTTGTTGGTATTATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 199
```

RESULT 11
US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Forsyth, R. Allyn
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001C1
; CURRENT APPLICATION NUMBER: US/10/771,241
; CURRENT FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
; LENGTH: 249
; TYPE: DNA
; ORGANISM: E. Coli
US-10-771-241-117

Query Match 6.4%; Score 29.2; DB 6; Length 249;
Best Local Similarity 52.5%; Pred. No. 1.8;
Matches 64; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||      || || ||| |    | |  | |    || | | |||
Db        180 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 121

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    |  |  ||  || || || ||| |  | |||    |||| | | |    || |
Db        120 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 61

Qy        415 CA 416
              |
Db         60 GA 59
```

RESULT 12
US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Forsyth, R. Allyn
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001C1
; CURRENT APPLICATION NUMBER: US/10/771,241
; CURRENT FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 757
; TYPE: DNA
; ORGANISM: E. Coli
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(757)
; OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

Query Match 6.4%; Score 29.2; DB 6; Length 757;
Best Local Similarity 52.5%; Pred. No. 3.1;
Matches 64; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||      || || ||| |    | |  | |    || | | |||
Db        264 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 323

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    |  |  ||  || || || ||| |  | |||    |||| | | |    || |
Db        324 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 383

Qy        415 CA 416
              |
Db        384 GA 385
```

RESULT 13
US-10-417-375A-143/c
; Sequence 143, Application US/10417375A
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375A
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143
; LENGTH: 83009
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(83009)
; OTHER INFORMATION: n = A,T,C or G
US-10-417-375A-143

Query Match 6.3%; Score 28.8; DB 6; Length 83009;
Best Local Similarity 48.8%; Pred. No. 38;
Matches 78; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```
Qy        281 AGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTA 340
              |||||  |   ||||   ||   | | || |  ||   |  |   | | |||  ||||
Db      14056 AGAATATTAAAAGAATGCATAATATCCTCAAAAAGAACATTGACTAGTATTCCTTTATTC 13997

Qy        341 CTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCA 400
              || ||    |      | ||||   |   |      || ||| |  |   |  ||| |
Db      13996 CTTTCAAGTACACTTAAGTTAATGCTTCAATGAAATACAAGCAGGCACCTATTACTTCTA 13937

Qy        401 GCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCAC 440
              |  |||||   | |||| | | ||   |  || | |||||
Db      13936 GAATAATGAGACCTCAGATGGATTACAGGCACGATGCCAC 13897
```

RESULT 14
PCT-US04-05654-2123
; Sequence 2123, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A

```
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Dubell III, Arnold N
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Repetti, Peter
;  APPLICANT:  Century, Karen
;  APPLICANT:  Gutterson, Neal
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Kumimoto, Roderick W
;  APPLICANT:  Pilgrim, Marsha L
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 PCT
;  CURRENT APPLICATION NUMBER: PCT/US04/05654
;  CURRENT FILING DATE:  2004-03-03
;  PRIOR APPLICATION NUMBER: 10/374,780
;  PRIOR FILING DATE: 2003-02-25
;  PRIOR APPLICATION NUMBER: 10/675,852
;  PRIOR FILING DATE: 2003-09-30
;  NUMBER OF SEQ ID NOS: 2950
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2123
;  LENGTH: 2511
;  TYPE: DNA
;  ORGANISM: Arabidopsis thaliana
;  FEATURE:
;  OTHER INFORMATION: G1548  Predicted polypeptide sequence is paralogous to G390, G391
PCT-US04-05654-2123

  Query Match             6.2%;  Score 28.4;  DB 1;  Length 2511;
  Best Local Similarity   56.4%;  Pred. No. 10;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         21 AGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGT 80
              ||||   ||||  |  ||||  || | |||||  |  | ||| |  | ||     || |||
Db       1470 AGCAGCAGCAGTTAAAGTAGGGCCTTGTAGTGCCCGAGTTGGAGGATTTGGAGGGCAGGT 1529

Qy         81 TACCCGTGTTGTTACCCATGAAATGGCACATGCA 114
              ||  | |     || | |||   || |  |||| |
Db       1530 TATACTTCCACTTGCTCATACTATTGAGCATGAA 1563


RESULT 15
US-10-021-698A-869/c
; Sequence 869, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 869
;  LENGTH: 20043
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (6103)
;  OTHER INFORMATION: a, t, c or g
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (20008)
;  OTHER INFORMATION: a, t, c or g
US-10-021-698A-869

  Query Match             6.2%;  Score 28.4;  DB 6;  Length 20043;
  Best Local Similarity   56.4%;  Pred. No. 27;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          9 TTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCTA 68
              | ||| | ||  |   | ||| || || || |  || ||| || | | | |||| | ||
Db      10361 TCTAATACTGTAATACTGCGCTGCTATTGTTGGCTCCGGCTGTCCACAGCCTGGGGGCTG 10302

Qy         69 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAA 102
              ||  ||   | | || |||    | | |  |||
Db      10301 GGAAGAGAGTGCTGCCTGTGGAAATGCTCGGGAA 10268


Search completed: March 19, 2004, 03:10:14
Job time : 57.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-15
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	223	48.9	680	12	BJ618688	BJ618688 BJ618688
c	2	37.4	8.2	484	9	AV430994	AV430994 AV430994
c	3	37.2	8.2	772	29	CG862779	CG862779 ZMMBBc027
c	4	35.8	7.9	484	14	CF089570	CF089570 QHM23E04.
	5	35.2	7.7	813	28	BZ204853	BZ204853 CH230-390
	6	34.6	7.6	702	12	BG441624	BG441624 GA__Ea001
	7	34.6	7.6	937	12	BG444255	BG444255 GA__Ea002
	8	34.2	7.5	296	9	AV058630	AV058630 AV058630
	9	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
c	10	34	7.5	811	28	BZ469000	BZ469000 BOOAO58TF
c	11	33.8	7.4	480	14	CF372418	CF372418 CSECS052B
c	12	33.8	7.4	481	14	CF414539	CF414539 CSECS071B
	13	33.8	7.4	509	14	CF605114	CF605114 RADIC01_0
	14	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	15	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
c	16	33.6	7.4	712	14	CF437982	CF437982 EST674327
	17	33.4	7.3	679	29	CC961526	CC961526 BOIGS10TF
c	18	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
c	19	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF
c	20	33.4	7.3	709	28	BH690914	BH690914 BOHXR64TF
c	21	33.4	7.3	785	28	BH578793	BH578793 BOGLY20TR
c	22	33.4	7.3	815	28	BH461716	BH461716 BOHIV03TR
c	23	33.4	7.3	827	28	BH438448	BH438448 BOGVQ28TR
c	24	33.4	7.3	830	28	BH715216	BH715216 BOMIM42TR
c	25	33.4	7.3	835	28	BH575194	BH575194 BOHJO27TF
c	26	33.4	7.3	853	28	BH578402	BH578402 BOHNV61TF
c	27	33.4	7.3	1026	28	BZ458784	BZ458784 BONHB14TF
	28	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
	29	33.2	7.3	625	10	BF501499	BF501499 AT16857.5
	30	33.2	7.3	788	9	AU139321	AU139321 AU139321
	31	33	7.2	456	12	BJ331339	BJ331339 BJ331339
	32	33	7.2	524	12	BJ367372	BJ367372 BJ367372
	33	33	7.2	547	12	BJ333958	BJ333958 BJ333958
	34	33	7.2	553	12	BJ336903	BJ336903 BJ336903
	35	33	7.2	569	12	BJ334047	BJ334047 BJ334047
	36	33	7.2	574	12	BJ330328	BJ330328 BJ330328
	37	33	7.2	607	12	BJ365766	BJ365766 BJ365766
	38	33	7.2	625	12	BJ396922	BJ396922 BJ396922
	39	33	7.2	629	12	BJ335378	BJ335378 BJ335378
	40	33	7.2	635	12	BJ397391	BJ397391 BJ397391
	41	33	7.2	645	12	BJ392753	BJ392753 BJ392753
	42	33	7.2	670	12	BJ336637	BJ336637 BJ336637
c	43	32.8	7.2	518	10	BF755518	BF755518 MR2-CT059
	44	32.8	7.2	572	12	BI626794	BI626794 RH67848.5
	45	32.8	7.2	574	10	BF854273	BF854273 MR2-EN009

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.

```
FEATURES        http://xenopus.nibb.ac.jp.
     source             Location/Qualifiers
                        1. .680
                        /organism="Xenopus laevis"
                        /mol_type="mRNA"
                        /db_xref="taxon:8355"
                        /clone="XL186b22"
                        /tissue_type="whole embryo"
                        /dev_stage="stage 10.5"
                        /clone_lib="NIBB Mochii normalized Xenopus early gastrula
ORIGIN                  library"

  Query Match                   48.9%;  Score 223;  DB 12;  Length 680;
  Best Local Similarity   68.0%;  Pred. No. 3.1e-58;
  Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120
              || ||   |  ||||    |  | | | |    |  ||    |  ||  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||  ||| |||||||||  ||  ||||| || || ||| |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  || || ||||| ||  ||| ||| | | | |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||| ||||| || || || || ||  || | | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| ||| |  ||| |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||||||||| || ||| | |||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138


RESULT 2
AV430994/c
LOCUS       AV430994                484 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PL027e10_r 5', mRNA sequence.
ACCESSION   AV430994
VERSION     AV430994.1  GI:8586219
KEYWORDS    EST.
SOURCE      Porphyra yezoensis
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
  PUBMED    10907854
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES        Location/Qualifiers
     source     1. .484
                /organism="Porphyra yezoensis"
                /mol_type="mRNA"
                /strain="TU-1"
                /db_xref="taxon:2788"
                /clone="PL027e10_r"
                /clone_lib="Porphyra yezoensis TU-1"
                /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
ORIGIN          XhoI"

  Query Match                   8.2%;  Score 37.4;  DB 9;  Length 484;
  Best Local Similarity   57.1%;  Pred. No. 2.4;
  Matches   68;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |  | ||  | | ||  ||| |||||  |||||  |  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
              | ||  ||   ||  |||  ||||  |||||||     ||      ||| | || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323


RESULT 3
CG862779/c
LOCUS       CG862779                772 bp    DNA     linear   GSS 19-NOV-2003
DEFINITION  ZMMBBc0273B12f ZMMBBc (EcoRI) Zea mays subsp. mays genomic clone
            ZMMBBc0273B12 5', genomic survey sequence.
ACCESSION   CG862779
VERSION     CG862779.1  GI:38436806
KEYWORDS    GSS.
SOURCE      Zea mays subsp. mays (maize)
  ORGANISM  Zea mays subsp. mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003c)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 450.
FEATURES        Location/Qualifiers
     source     1. .772
                /organism="Zea mays subsp. mays"
                /mol_type="genomic DNA"
                /cultivar="B73"
                /sub_species="mays"
                /db_xref="taxon:4578"
                /clone="ZMMBBc0273B12"
                /lab_host="E. coli DH10B"
                /clone_lib="ZMMBBc (EcoRI)"
                /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"
ORIGIN

  Query Match                   8.2%;  Score 37.2;  DB 29;  Length 772;
  Best Local Similarity   51.9%;  Pred. No. 3.6;
  Matches   84;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;
```

```
Qy         45 TGGCAGTGCTCTGGCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAAT 104
Db        202 TGTCCATCAGCTCGCTCGCCCTGGTGGGGGTCTTGCGTCCCAACTTGCTAACCAGGTCGC 143

Qy        105 GGCACATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGC 164
Db        142 GGCAGGTGGTTCCGGCAAGGACACGCCGATGAGCATCTGTCGGTATGTTGGGCAGCTCGG 83

Qy        165 TGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGAC 206
Db         82 TGAGCTGCATTTGAACATTCAGCAATGTAGTCCAGGAGAGAC 41
```

RESULT 4
CF089570/c

```
LOCUS       CF089570                484 bp    mRNA    linear   EST 22-JUL-2003
DEFINITION  QHM23E04.yg.ab1 QH_M sunflower H.argophyllus Helianthus argophyllus
            cDNA clone QHM23E04, mRNA sequence.
ACCESSION   CF089570
VERSION     CF089570.1  GI:33128637
KEYWORDS    EST.
SOURCE      Helianthus argophyllus
  ORGANISM  Helianthus argophyllus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QH_CA_Contig2049, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QHM23  row: E  column: 04.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Helianthus argophyllus"
                     /mol_type="mRNA"
                     /db_xref="taxon:73275"
                     /clone="QHM23E04"
                     /lab_host="E.coli"
                     /clone_lib="QH_M sunflower H.argophyllus"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from three different sources (seedling, root and leaf) of
                     RNA from a single genotype. cDNAs were pooled and
                     directionally cloned into a custom medium-copy vector.
                     Details of library construction can be obtained at
                     http://cgpdb.ucdavis.edu/"
ORIGIN
```

```
  Query Match             7.9%;  Score 35.8;  DB 14;  Length 484;
  Best Local Similarity   46.0%;  Pred. No. 7.4;
  Matches  121;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy         58 GCTGGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCC 117
Db        323 GCTGGGTTCATTGATCCTCCTGATACCGGTCCTGCTAAGATGGAAGTAATGCATACGGCT 264

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
Db        263 GATCCTACTGCCATACCGGCTAGCTCTCCTACGGCTTTAGAATCTGTAGCTACGGCTGAA 204

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
Db        203 GTGACAAACATCATTGAGAAAGTTACAATGATTTCCATGATCAGAGCTTGTAACTCTGTC 144

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
Db        143 CCAGAAGGCGTGGTTGTGCCTAGATTTGTTATTGGATCTAACATTATCCGTAGCGCAAAT 84

Qy        298 AATGCCACCATCGACCAGTGGAA 320
Db         83 GAAGCCAAGATGGACCCTGTGAA 61
```

RESULT 5
BZ204853

```
LOCUS       BZ204853                813 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-390D21, genomic survey sequence.
ACCESSION   BZ204853
VERSION     BZ204853.1  GI:23862905
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-390D21.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 390  row: D  column: 21
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-390D21"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN
```

```
  Query Match             7.7%;  Score 35.2;  DB 28;  Length 813;
  Best Local Similarity   49.0%;  Pred. No. 16;
  Matches   94;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy        178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
Db        548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
```

```
Db        608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357

Db        668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAAGGACAC 727

Qy        358 GGTAATAACGCC 369

Db        728 TGTGGTTAGGAC 739
```

RESULT 6
BG441624

```
LOCUS       BG441624                 702 bp    mRNA    linear   EST 15-MAR-2001
DEFINITION  GA__Ea0014A14f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0014A14f, mRNA sequence.
ACCESSION   BG441624
VERSION     BG441624.1  GI:13351276
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 696.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0014A14f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN
```

```
  Query Match             7.6%;  Score 34.6;  DB 12;  Length 702;
  Best Local Similarity   46.5%;  Pred. No. 22;
  Matches  112;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        185 AACAATGACGGGAATGCGAATGGGAATGTTGGCAGCACAGACAATCAGAGCAACACCAAT 244

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        245 GACAATGTAAATGCAAATGAAGATGTTACTCAAAAGGAAAACAACGATTCAAGTTCAAAT 304

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        305 GAAAATGCTGCACAGAGCAACACCAACAACAATGAAAACAATAACAATGAAAATGCGGGT 364

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        365 CAGGACGGGAAGAATGCCACTCAGAGTAACACTGACAACAATGAAAATGTGGGTCAGAAT 424

Qy        421 G 421

Db        425 G 425
```

RESULT 7
BG444255

```
LOCUS       BG444255                 937 bp    mRNA    linear   EST 15-MAR-2001
DEFINITION  GA__Ea0023M14f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0023M14f, mRNA sequence.
ACCESSION   BG444255
VERSION     BG444255.1  GI:13353907
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 937)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 804.
FEATURES             Location/Qualifiers
     source          1. .937
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0023M14f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN
```

```
  Query Match             7.6%;  Score 34.6;  DB 12;  Length 937;
  Best Local Similarity   46.5%;  Pred. No. 26;
  Matches  112;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        185 AACAATGACGGGAATGCGAATGGGAATGTTGGCAGCACAGACAATCAGAGCAACACCAAT 244

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        245 GACAATGTAAATGCAAATGAAGATGTTACTCAAAAGGAAAACAACGATTCAAGTTCAAAT 304

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        305 GAAAATGCTGCACAGAGCAACACCAACAACAATGAAAACAATAACAATGAAAATGCGGGT 364

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        365 CAGGACGGGAAGAATGCCACTCAGAGTAACACTGACAACAATGAAAATGTGGGTCAGAAT 424

Qy        421 G 421

Db        425 G 425
```

```
RESULT 8
AV058630
LOCUS       AV058630                296 bp    mRNA    linear   EST 23-JUN-1999
DEFINITION  AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA
            clone 1810054H02, mRNA sequence.
ACCESSION   AV058630
VERSION     AV058630.1  GI:5158377
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 296)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .296
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1810054H02"
                     /sex="male"
                     /tissue_type="pancreas"
                     /dev_stage="adult"
                     /clone_lib="Mus musculus pancreas C57BL/6J adult"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 9;  Length 296;
  Best Local Similarity   53.3%;  Pred. No. 17;
  Matches   72;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCG 310
              |||  ||  | |||  | |  |  | ||  ||  ||| |  |||||  | ||  | |  |
Db         13 GTGGAGACCACAGTTTTTTCTAGGTCACCAAGGCTGGGTAAAGAAAAGAAGCATCAACTG 72

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              |||||| | | | |||||    ||    | ||    || | | |  || | ||| |  |
Db         73 TCCAGTGAAGCACAAAAAATGAAGAAGAAAATGATTACCCAAAAGTTGGAATTAAGGTAG 132

Qy        371 CGCTGGTTAATCAGA 385
              | ||   ||  |||
Db        133 AGATGCCAAAATAGA 147


RESULT 9
BJ334624
LOCUS       BJ334624                523 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda47o06 5', mRNA sequence.
ACCESSION   BJ334624
VERSION     BJ334624.1  GI:19164754
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda47o06"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 12;  Length 523;
  Best Local Similarity   57.1%;  Pred. No. 25;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        398 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGANGACAATAAT 457

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACT 342
              |||  || || | || |    ||   ||||||    ||||| |
Db        458 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAAT 502


RESULT 10
BZ469000/c
LOCUS       BZ469000                811 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BOOAO58TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOAO58,
            genomic survey sequence.
ACCESSION   BZ469000
VERSION     BZ469000.1  GI:26764546
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOOAO58TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .811
```

/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOOAO58"
/clone_lib="BO_1.6_2_KB_tot"
/note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match 7.5%; Score 34; DB 28; Length 811;
Best Local Similarity 48.9%; Pred. No. 37;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```
Qy        176 TGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCG 235
              ||  | || |     ||||| |||      |||||  |    |  | |||||  ||  ||
Db        610 TGTTAGGCAAGCATTGTAAAACTGTGTTTCCCATTTCATCTACTGTTTATGACAAATGCT 551

Qy        236 CCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
              |||  ||   ||   ||  | |   | | ||   |    || ||  | || ||||  | |
Db        550 TTGATCTAATTCACTCTGATGTTTGGACTGCTCCTTGCTTGTCTAGGGATAGTTTTAAAT 491

Qy        296 ATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACG 355
              |   || |||  ||     |    ||  | ||| || ||   || ||  |     |||  |
Db        490 ACTTTGTCACATTCATAAATGAAAAATCGAAATACACCTGGATAACACTTATTCAAACCA 431

Qy        356 GCGGTA 361
                | ||
Db        430 AAGATA 425
```

RESULT 11
CF372418/c
LOCUS CF372418 480 bp mRNA linear EST 27-AUG-2003
DEFINITION CSECS052B08_FLOn0012 CabSau Normalised Flower Stage 12 (FLOn0012) Vitis vinifera cDNA clone CSECS052B08 3', mRNA sequence.
ACCESSION CF372418
VERSION CF372418.1 GI:34319664
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 480)
AUTHORS Iocco,P., Hua,C., Davies,C. and Thomas,M.R.
TITLE Expressed sequence tags from the grapevine cultivar Cabernet Sauvignon
JOURNAL Unpublished (2003)
COMMENT Contact: Mark R. Thomas
CSIRO Plant Industry
CSIRO
PO Box 350, Glen Osmond, SA, 5064, Australia
Tel: 61 8 83038600
Fax: 61 8 83038601
Email: Mark.R.Thomas@csiro.au
Seq primer: CCCAGTCACGACGTTGTAAAACG (M13 Forward)
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .480
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSECS052B08"
/sex="Hermaphrodite"
/dev_stage="12 - modified E-L system"
/clone_lib="CabSau Normalised Flower Stage 12 (FLOn0012)"
/note="Organ: Inflorescence including flowers; Vector: pZL; Normalised cDNA library from immature inflorescences at stage 12 of the modified E-L system. Tissue collected from field grown plants. A description of the modified E-L system can be found in the paper by B. G. Coombe 'Adoption of a system for identifying grapevine growth stages' (1995) Aust. J. Grape and Wine Res. 1: 104-110."

ORIGIN

Query Match 7.4%; Score 33.8; DB 14; Length 480;
Best Local Similarity 49.7%; Pred. No. 31;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy         94 ACCCATGAAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              | | ||||  ||| || ||    |   |||   | ||| |       |       |- || |
Db        222 ATCAATGAGGTGGAACCTGACCCTTCCCCTCCCCCAAAGCAACTGTTGCGAGGGGCCGCT 163

Qy        154 TCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACC 213
              |||||  |||||||  ||||       ||||| | |   |   || |||    |||| ||
Db        162 GCCGCTGCCGCTGCGGCTGCTGCTAGCGAAGACACGAGGGCTAAAAGCCACATTTCC 103

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              |||| |  |   |    ||   |      ||| || |  |||| ||  | |
Db        102 CAGAAAGAATCAAGGCTTGGAAGCAGCAGTAGCCGGGATCAGGATGATTCTCC 50
```

RESULT 12
CF414539/c
LOCUS CF414539 481 bp mRNA linear EST 02-SEP-2003
DEFINITION CSECS071B09_POSn0036 CabSau Normalised Berry Postveraison Stage 36 (POSn0036) Vitis vinifera cDNA clone CSECS071B09 3', mRNA sequence.
ACCESSION CF414539
VERSION CF414539.1 GI:34415785
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 481)
AUTHORS Iocco,P., Hua,C., Davies,C. and Thomas,M.R.
TITLE Expressed sequence tags from the grapevine cultivar Cabernet Sauvignon
JOURNAL Unpublished (2003)
COMMENT Contact: Mark R. Thomas
CSIRO Plant Industry
CSIRO
PO Box 350, Glen Osmond, SA, 5064, Australia
Tel: 61 8 83038600
Fax: 61 8 83038601
Email: Mark.R.Thomas@csiro.au
Seq primer: CCCAGTCACGACGTTGTAAAACG (M13 Forward)
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .481
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSECS071B09"
/sex="Hermaphrodite"
/dev_stage="36 - modified E-L system"
/clone_lib="CabSau Normalised Berry Postveraison Stage 36 (POSn0036)"
/note="Organ: Fruit with seeds removed; Vector: pZL; A cDNA library from postveraison fruit at stage 36 of the modified E-L system. Tissue collected from field grown plants. A description of the modified E-L system can be found in the paper by B. G. Coombe 'Adoption of a system for identifying grapevine growth stages' (1995) Aust. J. Grape and Wine Res. 1: 104-110."

ORIGIN

Query Match 7.4%; Score 33.8; DB 14; Length 481;
Best Local Similarity 49.7%; Pred. No. 31;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy         94 ACCCATGAAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
```

```
Db        223 ATCAATGAGGTGGAACCTGACCCTTCCCCTCCCCCAAAGCAACTGTTGCGAGGGGCCGCT 164

Qy        154 TCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACC 213
              ||||| ||||||| |||| |||||  | |   |  || |||  |||| ||
Db        163 GCCGCTGCCGCTGCGGCTGCTGCTGCTAGCGAAGACACGAGGGCTAAAAGCCACATTTCC 104

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||||  |  |   |    ||   |       ||| || |  |||| ||  | |
Db        103 CAGAAAGAATCAAGGCTTGGAAGCAGCAGTAGCCGGGATCAGGATGATTCTCC 51


RESULT 13
CF605114
LOCUS       CF605114                509 bp    mRNA    linear   EST 30-SEP-2003
DEFINITION  RADIC01_000700 Grape Root pSPORT1 Library Vitis vinifera cDNA 5',
            mRNA sequence.
ACCESSION   CF605114
VERSION     CF605114.1  GI:37185761
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Moser,C., Segala,C., Fontana,P., Salakhudtinov,I., Gatto,P.,
            Pindo,M., Zyprian,E., Toepfer,R., Grando,M.S. and Velasco,R.
  TITLE     Expressed sequence tags from different organs of Vitis vinifera
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Moser C
            Laboratorio di Genetica Molecolare
            Istituto Agrario di San Michele all'Adige (IASMA)
            via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia
            Tel: 0039-0461-615314
            Fax: 0039-0461-650956
            Email: claudio.moser@ismaa.it
            The sequencing work has been funded by the 'Fondazione Cassa di
            Risparmio di Trento e Rovereto'
            High quality sequence stop: 509.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Pinot noir"
                     /db_xref="taxon:29760"
                     /sex="Hermaphrodite"
                     /dev_stage="young root (first year)"
                     /lab_host="DH10B"
                     /clone_lib="Grape Root pSPORT1 Library"
                     /note="Organ: root; Vector: pSPORT1; Site_1: NotI; Site_2:
                     SalI"
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 509;
  Best Local Similarity   49.7%;  Pred. No. 32;
  Matches   86;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         94 ACCCATGAAATGGCACATGCATCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGT 153
              | | ||||  ||| || ||   |  |||  | ||| |     |       |  || |
Db        295 ATCAATGAGGTGGAACCTGACCCTTCCCCTCCCCCAAAGCAACTGTTGCGAGGGGCCGCT 354

Qy        154 TCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACC 213
              |||||  |||||||  ||||      ||||| | |   |   || |||    |||| ||
Db        355 GCCGCTGCCGCTGCGGCTGCTGCTGCTAGCGAAGACACGAGGGCTAAAAGCCACATTTCC 414

Qy        214 CAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
              ||||  |  |   |    ||   |       ||| || |  |||| ||  | |
Db        415 CAGAAAGAATCAAGGCTTGGAAGCAGCAGTAGCCGGGATCAGGATGATTCTCC 467


RESULT 14
BJ335653
LOCUS       BJ335653                594 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ335653 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda51h06 5', mRNA sequence.
ACCESSION   BJ335653
VERSION     BJ335653.1  GI:19165783
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda51h06"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.4%;  Score 33.6;  DB 12;  Length 594;
  Best Local Similarity   57.1%;  Pred. No. 41;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        223 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 282

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACT 342
              |||  || || | || |    ||   ||||||     |||||| |
Db        283 AATAACAACAACAACAACAATAATAATAAAAATAANAATATTAAT 327


RESULT 15
BJ336964
LOCUS       BJ336964                653 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ336964 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda55e20 5', mRNA sequence.
ACCESSION   BJ336964
VERSION     BJ336964.1  GI:19167094
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .653
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
```

```
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda55e20"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             7.4%;  Score 33.6;  DB 12;  Length 653;
  Best Local Similarity   53.9%;  Pred. No. 43;
  Matches   69;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              |||| ||   | |    ||| |||||   ||||| | | | |   ||||    ||  |||
Db        187 GATGAAGATGATGAATTGGAGAATAGATATATTGGAATCAATAGAAATGAGGACAATAAT 246

Qy        298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
              |||  || || | || |    ||   ||||||     |||||| |   |   |  | |
Db        247 AATAACAACAACAACAACAATAATAATAAAAATAAAAATATTAATAAAGATGATGAAGAA 306

Qy        358 GGTAATAA 365
                |||  ||
Db        307 AATAAAAA 314


Search completed: March 18, 2004, 20:11:28
Job time : 2034.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	406.4	89.1	456	6	I44909	I44909 Sequence 58
2	406.4	89.1	2067	1	SEU43280	U43280 Salmonella
3	400	87.7	5103	1	STAJ2301	AJ002301 Salmonell
4	400	87.7	22411	1	AE008749	AE008749 Salmonell
5	398.4	87.4	254050	1	AL627269	AL627269 Salmonell
c 6	398.4	87.4	301983	1	AE016840	AE016840 Salmonell
7	392.6	86.1	1048	1	STAGFBA	AJ000514 Salmonell
8	282.2	61.9	361	6	I44908	I44908 Sequence 56
9	280	61.4	2889	1	CSP515700	AJ515700 Citrobact
10	245.2	53.8	2920	1	CFR515701	AJ515701 Citrobact
11	232	50.9	4680	1	ECCSGABDG	X90754 E.coli csgG
12	232	50.9	10346	1	AE000205	AE000205 Escherich
13	232	50.9	15047	1	D90741	D90741 Escherichia
14	230.4	50.5	456	6	AX814811	AX814811 Sequence
15	230.4	50.5	648	1	ECOCSGAA	L04979 Escherichia
16	220.6	48.4	306358	1	AE016759	AE016759 Escherich
17	219	48.0	1711	1	AF275733	AF275733 Escherich
18	219	48.0	10190	1	AE005315	AE005315 Escherich
19	219	48.0	327773	1	AP002554	AP002554 Escherich
20	192	42.1	2883	1	ESA515702	AJ515702 Enterobac
21	179.4	39.3	230	1	SEU53207	U53207 Salmonella
22	163.2	35.8	10370	1	AE015131	AE015131 Shigella
23	163.2	35.8	292504	1	AE016981	AE016981 Shigella
c 24	140.2	30.7	437	1	AF237726	AF237726 Shigella
25	93.2	20.4	19201	1	D90742	D90742 Escherichia
26	64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
27	48.2	10.6	78	6	AX814809	AX814809 Sequence
28	39.4	8.6	10999	1	AE002536	AE002536 Neisseria
29	39.4	8.6	349980	6	AX044034	AX044034 Sequence
c 30	37.8	8.3	349061	1	NMA2Z2491	AL162753 Neisseria
c 31	37	8.1	9955	1	AE002504	AE002504 Neisseria
c 32	37	8.1	349980	6	AX044033	AX044033 Sequence
c 33	36.8	8.1	126062	8	AP006135	AP006135 Lotus cor
34	36.6	8.0	110000	2	AC142602_0	AC142602 Macaca mu
35	36.6	8.0	305730	2	AC114347	AC114347 Rattus no
36	36.6	8.0	311823	2	AC095362	AC095362 Rattus no
37	36.4	8.0	10798	1	AE004604	AE004604 Pseudomon
38	36.4	8.0	110000	2	AC110929_4	Continuation (5 of
39	36.4	8.0	181996	2	AC108614	AC108614 Rattus no
c 40	36.4	8.0	212174	2	AC105896	AC105896 Rattus no
c 41	36.4	8.0	305724	2	AC128763	AC128763 Rattus no
42	36.2	7.9	1797	3	LEIMSP52A	L19563 Leishmania
43	36.2	7.9	1860	3	LIGP63GEN	Z83677 L.infantum
44	36.2	7.9	2104	3	LDO495002	AJ495002 Leishmani
45	36.2	7.9	2106	3	LDO495006	AJ495006 Leishmani

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match              89.1%;  Score 406.4;  DB 6;  Length 456;
  Best Local Similarity    93.2%;  Pred. No. 1.9e-103;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||     |   ||   |  | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match              89.1%;  Score 406.4;  DB 1;  Length 2067;
  Best Local Similarity    93.2%;  Pred. No. 2.1e-103;
```

```
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||||      |   ||  |    | | ||  | |    || | ||| ||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301                5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1.
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match           87.7%;  Score 400;  DB 1;  Length 5103;
  Best Local Similarity 92.3%;  Pred. No. 1.4e-101;
  Matches  421;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |     |  ||  |    | | ||  | |    || | ||| ||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
     RBS                complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene               complement(2580. .4088)
                        /gene="STM1128"
     CDS                complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS                complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene               complement(4413. .5099)
                        /gene="STM1129"
     CDS                complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
     RBS                complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene               5599. .6759
                        /gene="STM1130"
     CDS                5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene               6793. .7497
                        /gene="STM1131"
     RBS                6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGPYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match           87.7%;  Score 400;  DB 1;  Length 22411;
  Best Local Similarity  92.3%;  Pred. No. 1.6e-101;
  Matches  421;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |      |    ||   |     | ||  | |    || | ||| ||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269              254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                /note="PS00142 Neutral zinc metallopeptidases,
                zinc-binding region signature"
     gene       3001. .4011
                /gene="STY1079"
                /note="synonym: pyrD"
     CDS        3001. .4011
                /gene="STY1079"
                /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="dihydroorotate dehydrogenase"
                /protein_id="CAD08184.1"
                /db_xref="GI:16502233"
                /db_xref="GOA:Q8Z7S9"
                /db_xref="SWISS-PROT:Q8Z7S9"
                /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                HI"
     misc_feature    3136. .3990
                /gene="STY1079"
                /note="Pfam match to entry PF01180 DHOdehase,
                Dihydroorotate dehydrogenase, score 563.50, E-value
                1.4e-165"
     misc_feature    3238. .3297
                /gene="STY1079"
                /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                /gene="STY1079"
                /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene       4174. .4719
                /gene="STY1080"
     CDS        4174. .4719
                /gene="STY1080"
                /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                discrepency in position of translational start site."
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAD08185.1"
                /db_xref="GI:16502234"
                /db_xref="SPTREMBL:Q8Z7S8"
                /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                IMNDRLKPQVHCHSFSLEQAV"
     gene       complement(4716. .5825)
                /gene="STY1081"
     CDS        complement(4716. .5825)
                /gene="STY1081"
                /note="Orthologue of E. coli P75863; Fasta hit to P75863
                (369 aa), 86% identity in 367 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAD08186.1"
                /db_xref="GI:16502235"
                /db_xref="GOA:Q8Z7S7"
                /db_xref="SPTREMBL:Q8Z7S7"
                /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                /gene="STY1081"
                /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                iron-sulfur cluster binding domains, score 55.90, E-value
                8.5e-13"
     gene       5924. .8032
                /gene="STY1082"
     CDS        5924. .8032
                /gene="STY1082"
                /note="Similar to Haemophilus influenzae hypothetical
                protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                scores: E(): 0, 60.3% id in 710 aa
                Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                (702 aa), 92% identity in 702 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAD08187.1"
                /db_xref="GI:16502236"
                /db_xref="GOA:Q8Z7S6"
                /db_xref="SPTREMBL:Q8Z7S6"
                /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                /gene="STY1082"
                /note="Pfam match to entry PF01170 UPF0020,
                Uncharacterized protein family UPF0020, score 781.30,
                E-value 3.7e-231"
     misc_feature    6506. .6541
                /gene="STY1082"
                /note="PS01261 Uncharacterized protein family UPF0020
                signature"
     gene       8045. .9952
                /gene="STY1083"
     CDS        8045. .9952
                /gene="STY1083"
                /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                526 aa overlap
                Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                overlap"

  Query Match             87.4%;  Score 398.4;  DB 1;  Length 254050;
  Best Local Similarity   92.1%;  Pred. No. 5.2e-101;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |      |    ||   |     | | ||  | |    || | ||| ||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361


RESULT 6
AE016840/c
LOCUS       AE016840              301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             87.4%;  Score 398.4;  DB 1;  Length 301983;
Best Local Similarity   92.1%;  Pred. No. 5.3e-101;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |      |    ||   |     | | ||  | |    || | ||| ||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match           86.1%;  Score 392.6;  DB 1;  Length 1048;
  Best Local Similarity  91.4%;  Pred. No. 1.6e-99;
  Matches  416;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||| |     |    ||   |     | | ||  | |    || | ||| ||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||| |||||||||||||| |
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           61.9%;  Score 282.2;  DB 6;  Length 361;
  Best Local Similarity  90.1%;  Pred. No. 1.9e-68;
  Matches  302;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
              ||||          |    ||   |     | | ||  | |    || | ||| ||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           61.4%;  Score 280;  DB 1;  Length 2889;
  Best Local Similarity  77.4%;  Pred. No. 9e-68;
  Matches  353;  Conservative    0;  Mismatches  100;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |    ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||  |     | |    ||   |     |    | || |    || | ||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |||||  || |||||||||| |||||||||    ||||  || | |||||||||||| ||
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||   ||||| || || |||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || |||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
CFR515701
LOCUS       CFR515701              2920 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515701
VERSION     AJ515701.1  GI:31790495
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2920)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2920
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /isolate="Fec4"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:546"
                     /country="Germany"
     gene            complement(211. .861)
                     /gene="csgD"
     CDS             complement(211. .861)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56673.1"
                     /db_xref="GI:31790496"
                     /translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
                     RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
                     HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1624. .2082
                     /gene="csgB"
     CDS             1624. .2082
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56674.1"
                     /db_xref="GI:31790497"
                     /translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
                     NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
                     ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
     gene            2123. .2572
                     /gene="csgA"
     CDS             2123. .2572
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56675.1"
                     /db_xref="GI:31790498"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
                     IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
                     QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN
```

```
  Query Match             53.8%;  Score 245.2;  DB 1;  Length 2920;
  Best Local Similarity   72.8%;  Pred. No. 5.8e-58;
  Matches  332;  Conservative    0;  Mismatches  118;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| |||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              || |||| |   | |     |   ||    ||   |  |  |    ||| |||| ||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              || ||||| |||||||| ||    ||||| || ||  |||  ||| ||||| || | |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || |||||||||| ||| | ||||| |||   |||| |||| || |||  || | ||
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| ||||| |||||||| || |||||  | |||||||  ||   ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||| |||||||||||||| || ||||| |||| ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || |||||||| | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

```
RESULT 11
ECCSGABDG
LOCUS       ECCSGABDG                4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
```

```
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           50.9%;  Score 232;  DB 1;  Length 4680;
  Best Local Similarity   69.3%;  Pred. No. 3.1e-54;
  Matches  316;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  | | |    |||| ||| ||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||    ||| | ||||| |||||| | |||  ||| | |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 12
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
```

and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

```
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423
```

```
Query Match           50.9%;  Score 232;  DB 1;  Length 10346;
Best Local Similarity 69.3%;  Pred. No. 3.3e-54;
Matches  316;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |     ||  ||      | |  |  | |     |||| ||| ||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448
```

RESULT 13
D90741
LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
FEATURES Location/Qualifiers
source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."
CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER

```
                ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS             2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
CDS             3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
gene            complement(4117. .4950)
                /gene="csgG"
CDS             complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
gene            complement(4977. .5393)
                /gene="csgF"
CDS             complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene            complement(5418. .5807)
                /gene="csgE"
CDS             complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
gene            complement(5812. .6462)
                /gene="csgD"
CDS             complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene            7217. .7672
                /gene="csgB"
CDS             7217. .7672

  Query Match           50.9%;  Score 232;  DB 1;  Length 15047;
  Best Local Similarity 69.3%;  Pred. No. 3.4e-54;
  Matches  316;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    || ||     | |  |  | |.    |||| ||| ||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||    ||     ||||| || || |||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || || ||    || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 14
AX814811
LOCUS       AX814811               456 bp    DNA     linear   PAT 05-DEC-2003
DEFINITION  Sequence 15 from Patent WO03064446.
ACCESSION   AX814811
VERSION     AX814811.1  GI:39104001
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
  TITLE     Peptides
  JOURNAL   Patent: WO 03064446-A 15 07-AUG-2003;
            Hansa Medical Research Aktiebolag (SE)
FEATURES             Location/Qualifiers
     source          1. .456
```

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           50.5%;  Score 230.4;  DB 6;  Length 456;
  Best Local Similarity  69.1%;  Pred. No. 7.6e-54;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | ||||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match           50.5%;  Score 230.4;  DB 1;  Length 648;
  Best Local Similarity  69.1%;  Pred. No. 7.8e-54;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | ||||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 18, 2004, 15:02:51
Job time : 2312.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64625	Aac64625 AgfA::PT3
	2	406.4	89.1	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	406.4	89.1	456	2	AAT74142	Aat74142 Salmonell
	4	406.4	89.1	456	3	AAC64617	Aac64617 Salmonell
	5	363.2	79.6	456	3	AAC64626	Aac64626 AgfA::PT3
	6	355.2	77.9	456	3	AAC64628	Aac64628 AgfA::PT3
	7	355.2	77.9	456	3	AAC64622	Aac64622 AgfA::PT3
	8	355.2	77.9	456	3	AAC64629	Aac64629 AgfA::PT3
	9	355.2	77.9	456	3	AAC64623	Aac64623 AgfA::PT3
	10	353.6	77.5	456	3	AAC64624	Aac64624 AgfA::PT3
	11	352	77.2	456	3	AAC64630	Aac64630 AgfA::PT3
	12	352	77.2	456	3	AAC64627	Aac64627 AgfA::PT3
	13	348.8	76.5	456	3	AAC64631	Aac64631 AgfA::PT3
	14	282.2	61.9	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	282.2	61.9	361	2	AAT74141	Aat74141 Salmonell
	16	230.4	50.5	456	3	AAC64619	Aac64619 Escherich
	17	230.4	50.5	456	9	ACF36153	Acf36153 E. coli C
	18	218	47.8	646	2	AAQ62647	Aaq62647 Fibronect
	19	156	34.2	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
c	21	50	11.0	78	3	AAC64609	Aac64609 AgfA (SEF
	22	50	11.0	78	3	AAC64610	Aac64610 AgfA (SEF
	23	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	24	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	25	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	26	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
	27	41.2	9.0	78	3	AAC64606	Aac64606 SefA (SEF
	28	40.8	8.9	100	7	ACD68809	Acd68809 E. coli K
c	29	40.6	8.9	78	3	AAC64605	Aac64605 SefA (SEF
	30	39.4	8.6	47475	3	AAA81465	Aaa81465 N. mening
	31	39.4	8.6	110000	3	AAA81489_5	Continuation (6 of
	32	39.4	8.6	349980	3	AAF21612	Aaf21612 Neisseria
	33	39.2	8.6	1083	5	AAS76745	Aas76745 DNA encod
c	34	37	8.1	14652	3	AAA81482	Aaa81482 N. mening
c	35	37	8.1	110000	3	AAA81489_1	Continuation (2 of
c	36	37	8.1	349980	3	AAF21611	Aaf21611 Neisseria
c	37	36.4	8.0	2610	7	ACA42270	Aca42270 Prokaryot
c	38	36	7.9	2667	7	ACA45315	Aca45315 Prokaryot
c	39	35.8	7.9	954	7	ACA25910	Aca25910 Prokaryot
	40	35.6	7.8	2000	7	ADA71938	Ada71938 Rice gene
	41	35.4	7.8	3300	6	ABS67377	Abs67377 Neisseria
c	42	35.4	7.8	3411	5	AAS88526	Aas88526 DNA encod
c	43	35.4	7.8	3412	5	AAS89144	Aas89144 DNA encod
c	44	34.8	7.6	2000	7	ADA71938	Ada71938 Rice gene
	45	34.6	7.6	456	3	AAC64620	Aac64620 Escherich

ALIGNMENTS

RESULT 1
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.3e-135;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX

FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 2; Length 456;
Best Local Similarity 93.2%; Pred. No. 9.9e-120;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||  |  ||  |  | | ||  | |  || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 3
AAT74142

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             89.1%;  Score 406.4;  DB 2;  Length 456;
  Best Local Similarity   93.2%;  Pred. No. 9.9e-120;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 3; Length 456;
Best Local Similarity 93.2%; Pred. No. 9.9e-120;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||  |  ||  |  | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;

XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 79.6%; Score 363.2; DB 3; Length 456;
Best Local Similarity 87.3%; Pred. No. 6.9e-106;
Matches 398; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||  |  ||  |  | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||  |||    |  ||||||  | |    | |    ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
AAC64628
ID AAC64628 standard; DNA; 456 BP.
XX
AC AAC64628;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36352.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 77.9%; Score 355.2; DB 3; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.5e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||  |||     ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  | ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 7
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.5e-103;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |   ||  |     | | ||  | |    || | |||  |||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||| |     | | || | ||  |    | | | |  |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |   ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.5e-103;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |   ||  |     | | ||  | |    || | |      ||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||  |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 9
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match   77.9%; Score 355.2; DB 3; Length 456;
  Best Local Similarity   86.2%; Pred. No. 2.5e-103;
  Matches 393; Conservative   0; Mismatches   63; Indels   0; Gaps   0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                    ||    | ||||  ||          |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 8.2e-103;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||  ||    | | | |  |  | |  |   ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |  ||  |   | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.7e-102;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | |||   |      | | ||   |    |       ||| ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36351.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.7e-102;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||  | |       | || |    || | |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match          76.5%;  Score 348.8;  DB 3;  Length 456;
  Best Local Similarity  85.3%;  Pred. No. 2.8e-101;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||  |  ||  |  | | ||  | |  || | ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||   || |||  |    |                |  |  ||  ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             61.9%;  Score 282.2;  DB 2;  Length 361;
  Best Local Similarity   90.1%;  Pred. No. 5.6e-80;
  Matches  302;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
              ||||       |    ||   |      | | ||  | |    || | ||| |||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             61.9%;  Score 282.2;  DB 2;  Length 361;
  Best Local Similarity   90.1%;  Pred. No. 5.6e-80;
  Matches  302;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
              ||||       |    ||   |      | | ||  | |    || | ||| |||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 18, 2004, 09:11:59
```

Job time : 290.405 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	406.4	89.1	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	282.2	61.9	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	36.4	8.0	798	4	US-09-252-991A-13717	Sequence 13717, A
c	4	36.4	8.0	2748	4	US-09-252-991A-13820	Sequence 13820, A
	5	36.4	8.0	3720	4	US-09-252-991A-13627	Sequence 13627, A
	6	33.2	7.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	7	33.2	7.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	8	33	7.2	834	4	US-09-621-976-2574	Sequence 2574, Ap
	9	32.4	7.1	332	2	US-08-469-412A-8	Sequence 8, Appli
	10	32.4	7.1	332	3	US-09-021-715-8	Sequence 8, Appli
c	11	31.4	6.9	2007	4	US-09-643-597-153	Sequence 153, App
c	12	31.4	6.9	2007	4	US-09-480-884A-153	Sequence 153, App
c	13	31.4	6.9	2007	4	US-09-542-615A-153	Sequence 153, App
c	14	31.4	6.9	2007	4	US-09-606-421B-153	Sequence 153, App
c	15	31.4	6.9	2007	4	US-09-221-107-153	Sequence 153, App
c	16	31.4	6.9	2148	4	US-09-643-597-154	Sequence 154, App
c	17	31.4	6.9	2148	4	US-09-480-884A-154	Sequence 154, App
c	18	31.4	6.9	2148	4	US-09-542-615A-154	Sequence 154, App
c	19	31.4	6.9	2148	4	US-09-606-421B-154	Sequence 154, App
c	20	31.4	6.9	2148	4	US-09-221-107-154	Sequence 154, App
c	21	31.2	6.8	1212	4	US-09-252-991A-7452	Sequence 7452, Ap
	22	31.2	6.8	1425	4	US-09-252-991A-7188	Sequence 7188, Ap
c	23	31.2	6.8	1491	6	5486473-3	Patent No. 5486473
c	24	31.2	6.8	1692	4	US-09-252-991A-7513	Sequence 7513, Ap
c	25	31.2	6.8	2436	4	US-09-540-236-492	Sequence 492, App
c	26	31.2	6.8	65792	4	US-09-596-002-31	Sequence 31, Appl
	27	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
	28	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	29	30.4	6.7	552	4	US-09-252-991A-9800	Sequence 9800, Ap
c	30	30.4	6.7	894	4	US-09-434-354-1	Sequence 1, Appli
c	31	30.2	6.6	774	4	US-09-252-991A-11893	Sequence 11893, A
c	32	30.2	6.6	1083	4	US-09-252-991A-11930	Sequence 11930, A
	33	30.2	6.6	3323	2	US-08-422-699A-10	Sequence 10, Appl
	34	30.2	6.6	3323	2	US-08-422-706B-10	Sequence 10, Appl
c	35	29.8	6.5	1500	4	US-09-252-991A-6254	Sequence 6254, Ap
	36	29.8	6.5	1695	4	US-09-252-991A-6044	Sequence 6044, Ap
c	37	29.8	6.5	1747	4	US-09-566-921-66	Sequence 66, Appl
c	38	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
c	39	29.4	6.4	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	40	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	41	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	42	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	43	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
c	44	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, App
c	45	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, App

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 89.1%; Score 406.4; DB 1; Length 456;
Best Local Similarity 93.2%; Pred. No. 3.1e-124;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 61.9%; Score 282.2; DB 1; Length 361;
Best Local Similarity 90.1%; Pred. No. 2.5e-83;
Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 3
US-09-252-991A-13717/c
; Sequence 13717, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13717
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13717

Query Match 8.0%; Score 36.4; DB 4; Length 798;
Best Local Similarity 47.4%; Pred. No. 0.04;
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```
Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
Db        270 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 211

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
Db        210 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGCTCGACCAGGGTGGCGATGTTG 151

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
```

```
Db        150 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 91

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
              ||       || ||  | |  |||| | | | ||||   || ||||||
Db         90 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAAACCGGCCTTGACGTAGGC 41


RESULT 4
US-09-252-991A-13820/c
; Sequence 13820, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 13820
;   LENGTH: 2748
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13820

  Query Match             8.0%;  Score 36.4;  DB 4;  Length 2748;
  Best Local Similarity   47.4%;  Pred. No. 0.074;
  Matches  109;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | | ||  |||  |  |||| |    | |     || |  || ||  |||  | |
Db        548 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 489

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              |||||||| | |   ||   ||||| || | |       || || ||      |||
Db        488 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGCTCGACCAGGGTGGCGATGTTG 429

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
              | |   |||    || | ||||  | | |       | |||    ||||    |   |
Db        428 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 369

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
              ||      || ||  | |  |||| | |  | ||||   || ||||||
Db        368 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAAACCGGCCTTGACGTAGGC 319


RESULT 5
US-09-252-991A-13627
; Sequence 13627, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 13627
;   LENGTH: 3720
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13627

  Query Match             8.0%;  Score 36.4;  DB 4;  Length 3720;
  Best Local Similarity   47.4%;  Pred. No. 0.087;
  Matches  109;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | |  ||  |||  |  |||| |    | |     || |  || ||  |||  | |
Db       3080 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 3139

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              |||||||| | |   ||   ||||| || | |       || || ||      |||
Db       3140 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGCTCGACCAGGGTGGCGATGTTG 3199

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
              | |   |||    || | ||||  | | |       | |||    ||||    |   |
Db       3200 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 3259

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
              ||      || ||  | |  |||| | |  | ||||   || ||||||
Db       3260 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAAACCGGCCTTGACGTAGGC 3309


RESULT 6
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             7.3%;  Score 33.2;  DB 3;  Length 4403765;
  Best Local Similarity   44.6%;  Pred. No. 31;
  Matches  131;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              | || | || |||| ||       | ||| | |  |  |  || | || | || | | |
Db    1093443 GCGGTGCCGACGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGTCGGGTCGATCACCGAC 1093502

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
                 |     ||| |      | |    |  || ||      |  | | || ||| |
Db    1093503 ACCGTGTTGCTGTCGAAGTTGGTCACGAAGACCAGGCCGGTGACGGGGCTGACCGCCACC 1093562

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGG 256
              | |     |||| || || |  | | |    ||    ||| | || ||| ||     || | |
Db    1093563 CCGCTTGGACCGTTGCCGATGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGCGGGGTTG 1093622

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316
              || | |      ||       | | | |||  | | |    |||| ||   | |
Db    1093623 ATCACCGACACCGTGCCGCTGCCGAAATTGGTGACGTAGACGTTGCCGCCCGGGTTGACC 1093682

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              |  || |        || |    ||| ||      | || ||| |   |||||
Db    1093683 GCCACCCCGTGCGGATCGTTGAAGCTGGCGTGGGTGATGGTGGTGACGGCGCCG 1093736


RESULT 7
```

```
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             7.3%;  Score 33.2;  DB 3;  Length 4411529;
  Best Local Similarity   44.6%;  Pred. No. 31;
  Matches  131;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              | || | || |||| ||      | ||| | | |   |  || | || | || | | |
Db    1093444 GCGGTGCCGACGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGTCGGGTCGATCACCGAC 1093503

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
              |     ||| |      | |   |  || ||     |  | | || ||| |
Db    1093504 ACCGTGTTGCTGTCGAAGTTGGTCACGAAGACCAGGCCGGTGACGGGGCTGACCGCCACC 1093563

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGG 256
              | |    |||| || || |  | |   ||  ||| | || ||| ||    || | |
Db    1093564 CCGCTTGGACCGTTGCCGATGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGCGGGGTTG 1093623

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316
              || |  |      ||        | | | ||||  |  | |   |||| ||   |  |
Db    1093624 ATCACCGACACCGTGCCGCTGCCGAAATTGGTGACGTAGACGTTGCCGCCCGGGTTGACC 1093683

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              |  || |       || |    ||| ||      | || ||| |   |||||
Db    1093684 GCCACCCCGTGCGGATCGTTGAAGCTGGCGTGGGTGATGGTGGTGACGGCGCCG 1093737


RESULT 8
US-09-621-976-2574
; Sequence 2574, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2574
;   LENGTH: 834
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 497..826
US-09-621-976-2574

  Query Match             7.2%;  Score 33;  DB 4;  Length 834;
  Best Local Similarity   14.9%;  Pred. No. 0.54;
  Matches   18;  Conservative   64;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         79 GGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTG 138
              :|  |: :: : :: :     ::| : :| |  :::| :: |: |::: : :::   ::
Db         35 RGGWGSCRSTSRMRRSKSYGARRCMSRRGKAKCMSWTGMSWCSTGRMSSWRKRKRSWSYW 94

Qy        139 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCT 198
              :|:: : |:::|  ::::  :::::: ::: :::::  |:::|:|:  |::|: |::|:
Db         95 STKWKMRGYSWTYSYRMYRCYSMWWWMCMRSCYKSRSRMCMMWGSGRSMCMYGWSARWCW 154

Qy        199 G 199
              :
Db        155 R 155


RESULT 9
US-08-469-412A-8
; Sequence 8, Application US/08469412A
; Patent No. 5856125
;  GENERAL INFORMATION:
;    APPLICANT:  Mavrothalassitis, George J.
;    APPLICANT:  Blair, Donald G.
;    APPLICANT:  Fisher, Robert J.
;    APPLICANT:  Beal Jr., Gregory J.
;    APPLICANT:  Athanasiou, Meropi A.
;    APPLICANT:  Sgouras, Dionyssios N.
;    TITLE OF INVENTION:  The ERF Genetic Locus and Its Products
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,412A
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  015280-229000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  332 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..332
;      OTHER INFORMATION:  /note= "human ERF gene promoter region"
US-08-469-412A-8

  Query Match             7.1%;  Score 32.4;  DB 2;  Length 332;
  Best Local Similarity   47.1%;  Pred. No. 0.53;
  Matches   99;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy         47 GCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCG 106
              || | |   ||| ||||  ||  | | |||| |  |||||||| |       |||| | |
Db        120 GCGGGGAGGGGGCGGGCGCAGTGTCTCCATGGCGACGCGGCGGTGACGTCGCCGGCCGGG 179

Qy        107 GCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGG 166
              |     | |  || | |     | | | | ||||| |  |   |    |  | ||
```

```
Db        180 GGGCGTGGGCGTCCCGGCCCCGGAGTGCGATATTAACCCGGGAGGCGGCGGCGGGGAGGG 239

Qy        167 CACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATG 226

Db        240 GAGAGGCTCTGAGAGGCGAGGCCGGGTGAGGCGGCGAGGGCGGCCCGACGGGCGCGGGAC 299

Qy        227 GTAACGGCGCCGATGTAGGCCAGGGTGCGG 256

Db        300 GGGACGGGGCAGCGAGGGCGCCGGGAGCCG 329


RESULT 10
US-09-021-715-8
; Sequence 8, Application US/09021715
; Patent No. 6194547
;   GENERAL INFORMATION:
;        APPLICANT: Mavrothalassitis, George J.
;                   Blair, Donald G.
;                   Fisher, Robert J.
;                   Beal Jr., Gregory J.
;                   Athanasiou, Meropi A.
;                   Sgouras, Dionyssios N.
;        TITLE OF INVENTION: The ERF Genetic Locus and Its Products
;        NUMBER OF SEQUENCES: 16
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/021,715
;             FILING DATE: 10-Feb-1998
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Garrett-Wackowski, Eugenia
;             REGISTRATION NUMBER: 37,330
;             REFERENCE/DOCKET NUMBER: 015280-229000
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 576-0200
;             TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 332 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  -
;             LOCATION:  1..332
;             OTHER INFORMATION:  /note= "human ERF gene promoter region"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-021-715-8

  Query Match             7.1%;  Score 32.4;  DB 3;  Length 332;
  Best Local Similarity   47.1%;  Pred. No. 0.53;
  Matches   99;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy         47 GCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCG 106

Db        120 GCGGGGAGGGGGCGGGCGCAGTGTCTCCATGGCGACGCGGCGGTGACGTCGCCGGCCGGG 179

Qy        107 GCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGG 166

Db        180 GGGCGTGGGCGTCCCGGCCCCGGAGTGCGATATTAACCCGGGAGGCGGCGGCGGGGAGGG 239

Qy        167 CACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATG 226

Db        240 GAGAGGCTCTGAGAGGCGAGGCCGGGTGAGGCGGCGAGGGCGGCCCGACGGGCGCGGGAC 299

Qy        227 GTAACGGCGCCGATGTAGGCCAGGGTGCGG 256

Db        300 GGGACGGGGCAGCGAGGGCGCCGGGAGCCG 329


RESULT 11
US-09-643-597-153/c
; Sequence 153, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.455C11
;  CURRENT APPLICATION NUMBER: US/09/643,597
;  CURRENT FILING DATE:  2000-08-21
;  NUMBER OF SEQ ID NOS:  369
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 153
;   LENGTH: 2007
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-643-597-153

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178

Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238

Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298

Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335

Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 12
US-09-480-884A-153/c
; Sequence 153, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C6
;  CURRENT APPLICATION NUMBER: US/09/480,884A
```

```
;  CURRENT FILING DATE:  2001-08-27
;  NUMBER OF SEQ ID NOS: 330
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 153
;    LENGTH: 2007
;    TYPE: DNA
;    ORGANISM: Homo sapien
US-09-480-884A-153

  Query Match                 6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              || | | | | | | | | | | | | | | | |
Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 13
US-09-542-615A-153/c
; Sequence 153, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Ligun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Fanger, Gary R.
; TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE:  210121.455C8
; CURRENT APPLICATION NUMBER: US/09/542,615A
; CURRENT FILING DATE:  2000-04-14
; NUMBER OF SEQ ID NOS:  350
; SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 153
;    LENGTH: 2007
;    TYPE: DNA
;    ORGANISM: Homo sapien
US-09-542-615A-153

  Query Match                 6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              || | | | | | | | | | | | | | | | |
Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 14
US-09-606-421B-153/c
; Sequence 153, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Fan, Ligun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C9
; CURRENT APPLICATION NUMBER: US/09/606,421B
; CURRENT FILING DATE:  2000-06-28
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 153
;    LENGTH: 2007
;    TYPE: DNA
;    ORGANISM: Homo sapien
US-09-606-421B-153

  Query Match                 6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              || | | | | | | | | | | | | | | | |
Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221


RESULT 15
US-09-221-107-153/c
; Sequence 153, Application US/09221107
; Patent No. 6660838
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C2
; CURRENT APPLICATION NUMBER: US/09/221,107
; CURRENT FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 153
;    LENGTH: 2007
;    TYPE: DNA
;    ORGANISM: Human
US-09-221-107-153

  Query Match                 6.9%;  Score 31.4;  DB 4;  Length 2007;
  Best Local Similarity   46.5%;  Pred. No. 2.8;
  Matches  101;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
```

```
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238

Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298

Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335

Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221
```

Search completed: March 18, 2004, 20:20:48
Job time : 64.4971 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	36.4	8.0	2610	12	US-10-282-122A-30140	Sequence 30140, A
c	2	36	7.9	2667	12	US-10-282-122A-33185	Sequence 33185, A
c	3	35.8	7.9	954	12	US-10-282-122A-13780	Sequence 13780, A
	4	34.2	7.5	1355	15	US-10-045-674-593	Sequence 593, App
	5	34.2	7.5	2547	12	US-10-282-122A-36154	Sequence 36154, A
c	6	34	7.5	1878	15	US-10-369-493-43934	Sequence 43934, A
	7	34	7.5	2886	9	US-09-801-368-131	Sequence 131, App
	8	33.6	7.4	2301	9	US-09-815-242-4844	Sequence 4844, Ap
	9	33.6	7.4	2301	9	US-09-815-242-9042	Sequence 9042, Ap
	10	33.4	7.3	3140	12	US-10-425-114-29450	Sequence 29450, A
	11	33.4	7.3	3485	12	US-10-424-599-64614	Sequence 64614, A
c	12	33.2	7.3	493	12	US-10-282-122A-26087	Sequence 26087, A
c	13	33.2	7.3	992	12	US-10-425-114-31747	Sequence 31747, A
	14	32.8	7.2	568	14	US-10-333-631-1	Sequence 1, Appli
	15	32.8	7.2	718	15	US-10-027-632-26874	Sequence 26874, A
	16	32.8	7.2	1059	9	US-09-738-626-482	Sequence 482, App
c	17	32.8	7.2	1113	9	US-09-738-626-481	Sequence 481, App
c	18	32.8	7.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
c	19	32.6	7.1	960	12	US-10-282-122A-11889	Sequence 11889, A
c	20	32.4	7.1	789	14	US-10-156-761-2022	Sequence 2022, Ap
c	21	32.4	7.1	9025608	14	US-10-156-761-1	Sequence 1, Appli
	22	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
	23	32.2	7.1	1015	14	US-10-198-846-4377	Sequence 4377, Ap
c	24	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
c	25	32.2	7.1	4830	12	US-10-282-122A-27869	Sequence 27869, A
	26	32	7.0	669	12	US-10-282-122A-31444	Sequence 31444, A
c	27	32	7.0	1001	12	US-10-425-114-14270	Sequence 14270, A
c	28	32	7.0	1152	14	US-10-156-761-303	Sequence 303, App
	29	32	7.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	30	31.8	7.0	678	14	US-10-156-761-1219	Sequence 1219, Ap
	31	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
	32	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
	33	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c	34	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
c	35	31.8	7.0	2766	12	US-10-282-122A-42042	Sequence 42042, A
	36	31.6	6.9	951	15	US-10-369-493-43201	Sequence 43201, A
c	37	31.6	6.9	1338	15	US-10-369-493-37494	Sequence 37494, A
c	38	31.6	6.9	1642	9	US-09-887-576-469	Sequence 469, App
	39	31.4	6.9	734	12	US-10-425-114-566	Sequence 566, App
c	40	31.4	6.9	1320	10	US-09-738-630-70	Sequence 70, Appl
c	41	31.4	6.9	1467	12	US-10-282-122A-25701	Sequence 25701, A
c	42	31.4	6.9	1833	9	US-09-989-920-87	Sequence 87, Appl
c	43	31.4	6.9	2007	9	US-09-735-705-153	Sequence 153, App
c	44	31.4	6.9	2007	9	US-09-850-716A-153	Sequence 153, App
c	45	31.4	6.9	2007	9	US-09-897-778-153	Sequence 153, App

ALIGNMENTS

RESULT 1
US-10-282-122A-30140/c
; Sequence 30140, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30140
; LENGTH: 2610
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30140

Query Match 8.0%; Score 36.4; DB 12; Length 2610;
Best Local Similarity 47.4%; Pred. No. 0.21;
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```
Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | |  ||  |||  |   |||| |     | |      ||| |  || ||  |||  | |
Db        410 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 351

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
                ||||||||  | |   ||   ||||| || | |     | || || ||      |||
Db        350 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGTTCGACCAGGGTGGCGATGTTG 291

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
              | |   |||    || | ||||  | | |       |  |||    ||||    |   |
Db        290 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 231

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
               ||      || ||  | |  ||||  | |    ||||   || ||||||
Db        230 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAATCCGGCCTTGACGTAGGC 181
```

RESULT 2
US-10-282-122A-33185/c
; Sequence 33185, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33185
; LENGTH: 2667
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-10-282-122A-33185

Query Match 7.9%; Score 36; DB 12; Length 2667;
Best Local Similarity 43.7%; Pred. No. 0.3;
Matches 159; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

```
Qy         25 TTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGC 84
              ||||  |  |||| |       ||    || |    | ||  ||      |  ||||||
Db        471 TTCGGCGACGTCGTGGAAGGCATCAAACATCAGCAGAGTGTGCTTGAGTTGTTCGGCGGC 412

Qy         85 GGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACC 144
              | | |    ||    |||| || | |    | || || |    | |||    | |   ||
Db        411 GACCGGGGCCAGCGTGGCGTCGTCGAGCAGTTCGACCAGTGTGACGATGTTGTAGCCGCC 352

Qy        145 CGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACG 204
              |    || | ||||  | | |    || ||     ||||    | | |  ||
Db        351 CTGCATGGTGCCCAGCAGCTCGGTTGCGCGCTTTTTGTCGATCAGGGGAGAAGTGGCTTC 292

Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              || ||  |||| || |  | || | ||||   || ||||||      |   | |   |
Db        291 GCCCTTGGCCAGGGCAGACAGGGAACCGGCCTTGACGTAGGCAGCTTCGTCAACACCCGG 232

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | ||   |   | | |  | |    ||   | |    | |  |  || | |||    |
Db        231 TGGAATGCGATTGGTGATCAGGTCGACAAGGAAAGCTTCTTCGCCAGCCGGGGGATTTTT 172

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAG 384
              | |   |||     |   |||    | ||||| |||     | ||  | ||  | |
Db        171 CAGCAGCTCGACCAGGCCTGCGGTTTGTTCGGCGTTAAGCGGCTGGGGCACGATACCCTG 112

Qy        385 ACCG 388
                ||
Db        111 GGCG 108
```

RESULT 3
US-10-282-122A-13780/c
; Sequence 13780, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

```
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13780
;   LENGTH: 954
;   TYPE: DNA
;   ORGANISM: Burkholderia mallei
US-10-282-122A-13780

  Query Match             7.9%;  Score 35.8;  DB 12;  Length 954;
  Best Local Similarity   54.1%;  Pred. No. 0.2;
  Matches   73;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy         55 CTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGT 114
              ||||| || | |     | ||| |||||||||| |||    ||||||  ||      |
Db        300 CTGGCGCGCATAGCCGAAGTAGGGGATGGCGGCGGTGATCCGGCCGGCGGATGCGCGCTT 241

Qy        115 TCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 174
                ||   ||| ||||||| | | |     | |||||     | |      |||| | |
Db        240 GAGCGCATCGACCATGATCATCAGTTCCATCAGGTTGTCGTTCGTCGGCGCGCACGTCGA 181

Qy        175 CTGCAAAGCGATGCC 189
               ||||   ||| | |
Db        180 TTGCAGGACGAAGAC 166


RESULT 4
US-10-045-674-593
; Sequence 593, Application US/10045674
; Publication No. US20030232333A1
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 60/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674-593

  Query Match             7.5%;  Score 34.2;  DB 15;  Length 1355;
  Best Local Similarity   26.9%;  Pred. No. 0.86;
  Matches   90;  Conservative   76;  Mismatches  169;  Indels    0;  Gaps    0;

Qy         20 CAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGG 79
              |: |  |:   ||   :|  | ::::||: |:| :   | :||:| :  :  :    | |
Db        755 CWCCKGTYAAYGCKGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWG 814

Qy         80 GCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGG 139
              |: ::|  ||:  :  :: :|| ||:||:: : |::::||:     : |: |: |:  :
Db        815 GYWSYGARGGWGGYGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARA 874

Qy        140 TTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTG 199
              |  |   :| :  :|     |  ||| |  |         |      |:|| |   |:::||
Db        875 TGGCWAAYGCYAAYAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTG 934

Qy        200 AAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259
              |  | |    :| :   : ::   | :|  | :  :| ||  | :   | :|    :|
Db        935 AYGCYAARGGYAARYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYA 994

Qy        260 ATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGA 319
               : |:  : |:        :   ||:||:  : |:  :| :|    : |:|     | :|
Db        995 TYGGYGAYGTYWSYGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKA 1054

Qy        320 ACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              |  |: |:|   |: |  |:  :| :||  |  |:
Db       1055 ATTCYCARATGGCYCARGTYGGWGAYGGKGAYAAY 1089


RESULT 5
US-10-282-122A-36154
; Sequence 36154, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
```

; SEQ ID NO 36154
; LENGTH: 2547
; TYPE: DNA
; ORGANISM: Streptococcus mutans
US-10-282-122A-36154

Query Match 7.5%; Score 34.2; DB 12; Length 2547;
Best Local Similarity 47.4%; Pred. No. 1.2;
Matches 102; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```
Qy        133 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGT 192
              |||| |||| | | ||||||| |||||||| |||| ||| |  |   |  ||
Db        850 CAGCAGGTTGCACTTGTTGTTGCCCATGAACTGGCGCATCAATGGTTTGGTAATCTGGTA 909

Qy        193 AAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGT 252
              |      |   |     |  ||    |     |||  |     ||| || |     |  |
Db        910 ACTATGAAGTGGTGGGATGACCTTTGGCTGAATGAAAGTTTTGCCAATATGATGGAATAT 969

Qy        253 GCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGAC 312
              | | ||  | ||||| |||  |    | |||  |||  | |   |    |||| ||
Db        970 GTTGCTATTGATGCTATTAAACCAAACTGGCATATTTTTGAAGACTTTCAGACCAGCGGT 1029

Qy        313 CAGTGGAACGCTAAAAACTCCGATATTACTGTCGG 347
                  | |||   ||    |||   |||| |||
Db       1030 GTTCCTCATGCTTTGAAGCGTGATGCAACTGACGG 1064
```

RESULT 6
US-10-369-493-43934/c
; Sequence 43934, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 43934
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: No. US20030233675A1toc punctiforme
US-10-369-493-43934

Query Match 7.5%; Score 34; DB 15; Length 1878;
Best Local Similarity 46.0%; Pred. No. 1.2;
Matches 115; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy         96 TAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTAC 155
              ||| |||| ||| |||| ||| || | | | | | | | | || |||||
Db        553 TAAAGGCGACGGAAATAACTCCTTCCTTGTGTCCGCGCCAAGGTTGAGTAATGGCGTTAC 494

Qy        156 CCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCA 215
              |    |||| |||| ||| |    ||          || |||| |    |      ||
Db        493 CTTGTAAATCCCACAAACACACCGTTCCGTCAAAACCAACACTGACAATGCTTACCCCAG 434

Qy        216 GAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACT 275
              | | |    |    | |||  |  | | || |  | ||||   | |      |
Db        433 AGGGGTTGGCGCAGCCATCGCTATTTGGGCTAAAGGCGACGGAAATAACTCCTTCCTCAT 374

Qy        276 GACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              | | | |    | || || ||| |    ||| | |         | |  |||| ||||
Db        373 GTCCCTGCCACGGTTGAGTAATGAGATTACCTTGTAGATCCCATAAACACAAAATTCCGT 314

Qy        336 TATTACTGTC 345
              | ||||| |
```

Db 313 CAATACTGCC 304

RESULT 7
US-09-801-368-131
; Sequence 131, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 131
; LENGTH: 2886
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-131

Query Match 7.5%; Score 34; DB 9; Length 2886;
Best Local Similarity 50.6%; Pred. No. 1.5;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              | ||  | | | | || | | | |||||    | ||    || | |   ||| | |
Db       1501 AACAACAACAATAACGATAACGATAACGATAACAATAATAGCAATAATAATAATAACAAT 1560

Qy        265 ACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCT 324
              | || | |    | | | |||  |  ||  ||||||  ||  | | || |    ||| |
Db       1561 AATAATAATAACAATAATAATAATAACAATAATAATAACAATAACAACAACAACAACAAT 1620

Qy        325 AAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAAC 366
              || ||     ||| || |     || |   || | |||||
Db       1621 AATAATAATAATAGTAATAATATTAAAAATAACGTTGATAAC 1662
```

RESULT 8
US-09-815-242-4844
; Sequence 4844, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

```
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4844
;   LENGTH: 2301
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-815-242-4844

  Query Match             7.4%;  Score 33.6;  DB 9;  Length 2301;
  Best Local Similarity   47.6%;  Pred. No. 1.9;
  Matches   99;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy        159 TGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAG 218
              |||||  | | |||||  ||     |||| | | |||| ||||      | |    | |
Db       1146 TGAAAGCGTATATGCAAAGCTTGATGATGTCAGCAAATTTGAAGATGTGACTGATAATAT 1205

Qy        219 CGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGAC 278
              |||  | |   |  | |||  ||  |   |    |  || || | | || |  |
Db       1206 GAGTTTAGATTTTGATACTAATGGTGGTTATTCTTTAAACTTTAATAATTTAGACCAAAG 1265

Qy        279 TCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATAT 338
              | | |||  | | | |||  |||         |  | |  || ||||  ||||  ||  |
Db       1266 TAAAAATTATGTAATAAAATATGAAGGGTATTATGATTCAAATGCTAGCAACTTAGAATT 1325

Qy        339 TACTGTCGGCCAATACGGCGGTAATAAC 366
              |       ||  |  ||   | |||||
Db       1326 TCAAACACACCTTTTTGGATATTATAAC 1353


RESULT 9
US-09-815-242-9042
; Sequence 9042, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9042
;   LENGTH: 2301
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2301)
US-09-815-242-9042

  Query Match             7.4%;  Score 33.6;  DB 9;  Length 2301;
  Best Local Similarity   47.6%;  Pred. No. 1.9;
  Matches   99;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy        159 TGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAG 218
              |||||  | | |||||  ||     |||| | | |||| ||||      | |    | |
Db       1146 TGAAAGCGTATATGCAAAGCTTGATGATGTCAGCAAATTTGAAGATGTGACTGATAATAT 1205

Qy        219 CGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGAC 278
              |||  | |   |  | |||  ||  |   |    |  || || | | || |  |
Db       1206 GAGTTTAGATTTTGATACTAATGGTGGTTATTCTTTAAACTTTAATAATTTAGACCAAAG 1265

Qy        279 TCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATAT 338
              | | |||  | | | |||  |||         |  | |  || ||||  ||||  ||  |
Db       1266 TAAAAATTATGTAATAAAATATGAAGGGTATTATGATTCAAATGCTAGCAACTTAGAATT 1325

Qy        339 TACTGTCGGCCAATACGGCGGTAATAAC 366
              |       ||  |  ||   | |||||
Db       1326 TCAAACACACCTTTTTGGATATTATAAC 1353


RESULT 10
US-10-425-114-29450
; Sequence 29450, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 29450
;   LENGTH: 3140
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-GMFLMINSOY044G01_FLI
US-10-425-114-29450

  Query Match             7.3%;  Score 33.4;  DB 12;  Length 3140;
  Best Local Similarity   45.8%;  Pred. No. 2.6;
  Matches  115;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy         91 AATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTT 150
              |||||   |   |||  |||||       | ||   |||||  |     |  |  ||
Db       1801 AATCAAGGCAAGGGCCTAAATAGGCTAAATCAGGTAGATGATAAAGGCATAGCTGGTAAA 1860

Qy        151 GTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATT 210
              | |   || |   ||    | | || | |  ||||| |  ||   ||||||  |||  |
Db       1861 ACTGCATCTGGAGGAGCTAGAGAAATGGACAAGGATGCTCAGAATGTTGAAACAAGTGGT 1920

Qy        211 ACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATT 270
              |  | ||  ||| ||  || ||      |||| |   |  |  |||||||  |    ||
Db       1921 TCAGATAGTAGTTCTGCAAAAGGAAAGAATGTTGTTGATAATATGGATAATGGAGAGCTT 1980
```

```
Qy        271 GAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAAC 330
Db       1981 TCAAAATCGAATGAGCGTCTTAAAAGAACTGCAGTTGAGGAAAATCCAGAGGATGAAAAG 2040

Qy        331 TCCGATATTAC 341
Db       2041 ATTGAACTTTC 2051
```

RESULT 11
US-10-424-599-64614
; Sequence 64614, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 64614
; LENGTH: 3485
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_29358C.1
US-10-424-599-64614

Query Match 7.3%; Score 33.4; DB 12; Length 3485;
Best Local Similarity 45.8%; Pred. No. 2.7;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy         91 AATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTT 150
Db       2146 AATCAAGGCAAGGGCCTAAATAGGCTAAATCAGGTAGATGATAAAGGCATACCTGGTAAA 2205

Qy        151 GTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATT 210
Db       2206 ACTGCATCTGGAGGAGCTAGAGAAATGGACAAGGATGCTCAGAATGTTGAAACAAGTGGT 2265

Qy        211 ACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATT 270
Db       2266 TCAGATAGTAGTTCTGCAAAAGGAAAGAATGTTGTTGATAATATGGATAATGGAGAGCTT 2325

Qy        271 GAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAAC 330
Db       2326 TCAAAATCAAATGAGCGTCTTAAAAGAACTGCAGTTGAGGAAAATCCAGAGGATGAAAAG 2385

Qy        331 TCCGATATTAC 341
Db       2386 ATTGAACTTTC 2396
```

RESULT 12
US-10-282-122A-26087/c
; Sequence 26087, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26087
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26087

Query Match 7.3%; Score 33.2; DB 12; Length 493;
Best Local Similarity 44.6%; Pred. No. 1.1;
Matches 131; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

```
Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
Db        408 GCGGTGCCGACGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGTCGGGTCGATCACCGAC 349

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
Db        348 ACCGTGTTGCTGTCGAAGTTGGTCACGAAGACCAGGCCGGTGACGGGGCTGACCGCCACC 289

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGG 256
Db        288 CCGCTTGGACCGTTGCCGATGGTGATGGGGGAGCCGGTGACGGTGTTGGTGGCGGGGTTG 229

Qy        257 ATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316
Db        228 ATCACCGACACCGTGCCGCTGCCGAAATTGGTGACGTAGACGTTGCCGCCCGGGTTGACC 169

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
Db        168 GCCACCCCGTGCGGATCGTTGAAGCTGGCGTGGGTGATGGTGGTGACGGCGCCG 115
```

RESULT 13
US-10-425-114-31747/c
; Sequence 31747, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 31747
;   LENGTH: 992
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMFLB73212F11_FLI
US-10-425-114-31747

  Query Match             7.3%;  Score 33.2;  DB 12;  Length 992;
  Best Local Similarity   52.1%;  Pred. No. 1.6;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         19 GCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGG 78
              || | |  || | ||  | | | |   ||    | ||| |  ||| | | |   |   |
Db        497 GCTGGAGGCGGATCAGGCCTTGGTGGCGGGTCAGGTCTCGGCGGCCTTGGTGGTCTCGGC 438

Qy         79 GGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTG 138
              ||||||||||   |     |||||||| | ||  |||| | |  |  | || |
Db        437 GGCGGCGGCGGCCTTGGCGGTGGCGGCGGCCTTGGTGGCGGCTCAGGATCGGGTCTGGGC 378

Qy        139 GTTACCCGTGTTGTTACCCATG 160
              |  | | |||   ||  |  ||
Db        377 GGCAGCGGTGGCCTTGGCGGTG 356


RESULT 14
US-10-333-631-1
; Sequence 1, Application US/10333631
; Publication No. US20030190651A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer AG
;  TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
;  TITLE OF INVENTION:  ENZYME
;  FILE REFERENCE: LIO121 Foreign Countries
;  CURRENT APPLICATION NUMBER: US/10/333,631
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US 60/221,917
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/280,112
;  PRIOR FILING DATE: 2001-04-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 568
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90)..(90)
;   OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

  Query Match             7.2%;  Score 32.8;  DB 14;  Length 568;
  Best Local Similarity   51.0%;  Pred. No. 1.6;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        139 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCT 198
              ||  |   || ||    |   ||||   |  | ||| |||||   ||   || ||     ||
Db         87 GTCNCGGCTGCTGCAGCTGCTGAACAAGGCCTTGCCCTGCATCACGCGCCCTGTCGTGCT 146

Qy        199 GAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGAT 258
              ||||  |   |  |  ||   ||  |    || || | || | || | ||| ||||||||
Db        147 GAAATGGTCCGTCACCCAGGCACTGGACGTCACAGAGCAGCTGGATGCCGGGGTGCGGTA 206

Qy        259 AATAGTACTATTGAACTGACTCAGAATGG 287
              | |  ||   ||    | ||   |||
Db        207 CCTGGATCTGCGGATAGCCCACATGCTGG 235


RESULT 15
US-10-027-632-26874
; Sequence 26874, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26874
;   LENGTH: 718
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-26874

  Query Match             7.2%;  Score 32.8;  DB 15;  Length 718;
  Best Local Similarity   49.4%;  Pred. No. 1.9;
  Matches   85;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         18 GGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATG 77
              || ||||| | |||||   |  |   || ||  |  | || | || || | | | ||  |
Db        514 GGAAGCATCCACAGCACAAGGTGCCGCTCGCCCTCGTGTGCCCGGGGTGGCTGCCCAGGG 573

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCT 137
              |   |  |         | |  ||| |  |||   |||||||   |         |
Db        574 GCTGGCAGCGTCCCCACGTGTCGGGGCTGGCCTGAGTTCCGTGTCCCTGGGATCTACAGC 633

Qy        138 GGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCC 189
              ||   || ||  |  | |||||   ||||| |||| |      | |||||
Db        634 TGTGTTCCCTGAAGGCAGACATGACTGGGCACAGGCACGGGGGTTCTATGCC 685


Search completed: March 19, 2004, 03:49:54
Job time : 264.734 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-17	Sequence 17, Appl
2	406.4	89.1	456	6	US-08-233-642A-56	Sequence 56, Appl
3	406.4	89.1	456	23	US-09-543-407-1	Sequence 1, Appli
4	363.2	79.6	456	23	US-09-543-407-19	Sequence 19, Appl
5	355.2	77.9	456	23	US-09-543-407-11	Sequence 11, Appl
6	355.2	77.9	456	23	US-09-543-407-13	Sequence 13, Appl
7	355.2	77.9	456	23	US-09-543-407-23	Sequence 23, Appl
8	355.2	77.9	456	23	US-09-543-407-25	Sequence 25, Appl
9	353.6	77.5	456	23	US-09-543-407-15	Sequence 15, Appl
10	352	77.2	456	23	US-09-543-407-21	Sequence 21, Appl
11	352	77.2	456	23	US-09-543-407-27	Sequence 27, Appl
12	348.8	76.5	456	23	US-09-543-407-29	Sequence 29, Appl
13	282.2	61.9	361	6	US-08-233-642A-54	Sequence 54, Appl
14	230.4	50.5	456	23	US-09-543-407-3	Sequence 3, Appli
15	230.4	50.5	675	13	US-08-978-878-3	Sequence 3, Appli
16	230.4	50.5	675	31	US-09-741-873B-3	Sequence 3, Appli
17	228	50.0	453	92	US-60-352-946-1	Sequence 1, Appli
18	228	50.0	453	101	US-60-444-371-1	Sequence 1, Appli
19	219.2	48.1	477	17	US-09-252-691-172	Sequence 172, App
20	219.2	48.1	477	17	US-09-252-691C-172	Sequence 172, App
21	219.2	48.1	477	51	US-10-417-886-172	Sequence 172, App
22	186.4	40.9	396	13	US-08-978-878-1	Sequence 1, Appli
23	161.6	35.4	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
c 28	50	11.0	78	23	US-09-543-407-52	Sequence 52, Appl
29	50	11.0	78	23	US-09-543-407-53	Sequence 53, Appl
30	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
31	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
32	41.2	9.0	78	23	US-09-543-407-49	Sequence 49, Appl
33	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
c 34	40.6	8.9	78	23	US-09-543-407-48	Sequence 48, Appl
c 35	39.4	8.6	4762	62	US-60-068-138-239	Sequence 239, App
36	39.4	8.6	47475	33	US-09-806-866A-13	Sequence 13, Appl
37	39.4	8.6	2242716	33	US-09-806-866A-1068	Sequence 1068, Ap
38	39.4	8.6	2272325	43	US-10-018-470A-1	Sequence 1, Appli
39	39.2	8.6	1083	1	PCT-US01-08631-12549	Sequence 12549, A
c 40	37	8.1	14652	33	US-09-806-866A-30	Sequence 30, Appl
c 41	37	8.1	2242716	33	US-09-806-866A-1068	Sequence 1068, Ap
c 42	37	8.1	2272325	43	US-10-018-470A-1	Sequence 1, Appli
43	36.4	8.0	561	17	US-09-252-691-171	Sequence 171, App
44	36.4	8.0	561	17	US-09-252-691C-171	Sequence 171, App
45	36.4	8.0	561	51	US-10-417-886-171	Sequence 171, App

ALIGNMENTS

```
RESULT 1
US-09-543-407-17
, Sequence 17, Application US/09543407
, GENERAL INFORMATION:
,  APPLICANT: White, Aaron P.
,  APPLICANT:  Doran, James L.
,  APPLICANT:  Collinson, S. Karen
,  APPLICANT:  Kay, William W.
,  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
,  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
,  FILE REFERENCE: 920043.406
,  CURRENT APPLICATION NUMBER: US/09/543,407
,  CURRENT FILING DATE:  2000-04-05
,  NUMBER OF SEQ ID NOS: 59
,  SOFTWARE: FastSEQ for Windows Version 4.0
, SEQ ID NO 17
,   LENGTH: 456
,   TYPE: DNA
,   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-133;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;       REGISTRATION NUMBER:  35,570
;       REFERENCE/DOCKET NUMBER:  920043.403C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANBERRY
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  456 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..456
US-08-233-642A-56

  Query Match             89.1%;  Score 406.4;  DB 6;  Length 456;
  Best Local Similarity   93.2%;  Pred. No. 1.2e-117;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||    |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             89.1%;  Score 406.4;  DB 23;  Length 456;
  Best Local Similarity   93.2%;  Pred. No. 1.2e-117;
  Matches  425;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||    |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             79.6%;  Score 363.2;  DB 23;  Length 456;
  Best Local Similarity   87.3%;  Pred. No. 6e-104;
  Matches  398;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |    | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||| |||    |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.1e-101;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |    | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
```

```
Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||| |          | | || | ||   |    | | | |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.1e-101;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |    | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||    | ||||  ||        |  || | ||||||||||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.1e-101;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||  |||      ||   ||      |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |  |  ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 2.1e-101;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |    ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              | | |||||     |          ||  | ||    ||||||||||||||||||||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             77.5%;  Score 353.6;  DB 23;  Length 456;
  Best Local Similarity   86.0%;  Pred. No. 6.7e-101;
  Matches  392;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||  ||   | | | |   | |     |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |   ||   |   | | ||  | |   || | |||  ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.2e-100;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              |||||||       |   ||   |   | | ||  | |   || | |||  ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||  |  |     | ||  |   || |  | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             77.2%;  Score 352;  DB 23;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 2.2e-100;
  Matches  391;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 76.5%; Score 348.8; DB 23; Length 456;
Best Local Similarity 85.3%; Pred. No. 2.2e-99;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||||||| |  ||  |  | | ||  | |  || | ||| ||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||  ||  |||  |   |                |   | |  ||   ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

Query Match 61.9%; Score 282.2; DB 6; Length 361;
Best Local Similarity 90.1%; Pred. No. 2.8e-78;
Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGC 183
              ||||  |  ||  |  | | ||  | |  || | ||| ||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              ||||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-3

  Query Match            50.5%;  Score 230.4;  DB 23;  Length 456;
  Best Local Similarity  69.1%;  Pred. No. 8.2e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | ||||||||||| | || || || |||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || |||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  | | |    |||| ||| ||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||||`|||||   ||| |||||||||  ||    ||||| || || |||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||  ||||| || || || || ||   || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||   ||| | ||||| |||||| | |||  ||| |  ||| |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
;  APPLICANT: NORMARK, Staffan
;  APPLICANT: OLSEN, Arne
;  TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
;  FILE REFERENCE: 012889-081
;  CURRENT APPLICATION NUMBER: US/08/978,878
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: SE 8801723-1
;  EARLIER FILING DATE: 1988-05-06
;  EARLIER APPLICATION NUMBER: US 07/347,189
;  EARLIER FILING DATE: 1989-05-04
;  EARLIER APPLICATION NUMBER: US 07/789,437
;  EARLIER FILING DATE: 1991-11-06
;  EARLIER APPLICATION NUMBER: US 07/970,846
;  EARLIER FILING DATE: 1992-11-03
;  EARLIER APPLICATION NUMBER: US 08/187,865
;  EARLIER FILING DATE: 1994-01-28
;  EARLIER APPLICATION NUMBER: US 08/318,519
;  EARLIER FILING DATE: 1994-10-05
;  EARLIER APPLICATION NUMBER: US 08/495,959
;  EARLIER FILING DATE: 1995-06-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-08-978-878-3

  Query Match            50.5%;  Score 230.4;  DB 13;  Length 675;
  Best Local Similarity  69.1%;  Pred. No. 9.4e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  ||||||||||||||| ||    || |||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  | | |    |||| ||| ||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||  ||    ||||| || || |||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || |||||||||  ||||| || || || || ||   || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||   ||| | ||||| |||||| | |||  ||| |  ||| |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

Search completed: March 19, 2004, 03:01:56
Job time : 2655.77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	230.4	50.5	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	186.4	40.9	396	5	US-09-741-873C-1	Sequence 1, Appli
	3	34.2	7.5	1355	6	US-10-045-674A-593	Sequence 593, App
c	4	31.4	6.9	435	6	US-10-021-698A-1216	Sequence 1216, Ap
c	5	31.4	6.9	2007	6	US-10-775-972-153	Sequence 153, App
c	6	31.4	6.9	2148	6	US-10-775-972-154	Sequence 154, App
c	7	31.2	6.8	9278	6	US-10-021-698A-796	Sequence 796, App
	8	31	6.8	1099	6	US-10-767-701-8827	Sequence 8827, Ap
	9	31	6.8	150223	6	US-10-624-149A-1	Sequence 1, Appli
c	10	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
c	11	30.4	6.7	883	6	US-10-767-701-1287	Sequence 1287, Ap
c	12	30.2	6.6	989	6	US-10-767-795-6329	Sequence 6329, Ap
c	13	30	6.6	1203	6	US-10-767-701-12486	Sequence 12486, A
c	14	29.8	6.5	1228	6	US-10-767-701-6233	Sequence 6233, Ap
c	15	29.6	6.5	656	6	US-10-767-701-20611	Sequence 20611, A
c	16	29.6	6.5	24900	6	US-10-767-471-10736	Sequence 10736, A
	17	29.6	6.5	186449	6	US-10-021-698A-715	Sequence 715, App
	18	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
c	19	29.4	6.4	302	6	US-10-767-701-17781	Sequence 17781, A
c	20	29.4	6.4	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	21	29.4	6.4	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	22	29.4	6.4	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	23	29.4	6.4	1400	7	US-60-545-213-6407	Sequence 6407, Ap
	24	29.4	6.4	145597	6	US-10-624-149A-2	Sequence 2, Appli
c	25	29.4	6.4	145597	6	US-10-624-149A-2	Sequence 2, Appli
c	26	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	27	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
	28	29	6.4	594	6	US-10-767-701-15041	Sequence 15041, A
c	29	29	6.4	1002	1	PCT-US04-05654-859	Sequence 859, App
	30	29	6.4	1052	6	US-10-767-701-12607	Sequence 12607, A
	31	29	6.4	2600	6	US-10-451-467A-39	Sequence 39, Appl
	32	28.8	6.3	959	6	US-10-767-701-1446	Sequence 1446, Ap
c	33	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	34	28.6	6.3	592	6	US-10-767-701-22979	Sequence 22979, A
	35	28.6	6.3	749	6	US-10-767-701-12516	Sequence 12516, A
	36	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
c	37	28.4	6.2	270	6	US-10-794-929-95	Sequence 95, Appl
	38	28.4	6.2	378	6	US-10-417-884A-1059	Sequence 1059, Ap
	39	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	40	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
c	41	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	42	28.2	6.2	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	43	28.2	6.2	3416	6	US-10-100-683-3397	Sequence 3397, Ap
	44	28.2	6.2	3776	7	US-60-548-091-34	Sequence 34, Appl
	45	28.2	6.2	3828	7	US-60-548-091-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 50.5%; Score 230.4; DB 5; Length 675;
Best Local Similarity 69.1%; Pred. No. 1.1e-68;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873C
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873C-1

  Query Match             40.9%;  Score 186.4;  DB 5;  Length 396;
  Best Local Similarity   66.9%;  Pred. No. 8.4e-54;
  Matches  265;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396


RESULT 3
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
;  APPLICANT: LADNER, ROBERT C.
;  APPLICANT:  COHEN, EDWARD H.
;  APPLICANT:  NASTRI, HORACIO G.
;  APPLICANT:  ROOKEY, KRISTIN L.
;  APPLICANT:  HOET, RENE
;  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
;  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
;  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
;  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
;  TITLE OF INVENTION:  LIBRARIES
;  FILE REFERENCE: DYAX/002 CIP2
;  CURRENT APPLICATION NUMBER: US/10/045,674A
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: 06/198,069
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: 09/837,306
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 635
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
;   LENGTH: 1355
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: M13-III
;   OTHER INFORMATION: nucleotide sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1305)
US-10-045-674A-593

  Query Match             7.5%;  Score 34.2;  DB 6;  Length 1355;
  Best Local Similarity   26.9%;  Pred. No. 0.098;
  Matches   90;  Conservative   76;  Mismatches  169;  Indels    0;  Gaps    0;

Qy         20 CAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGG 79

Db        755 CWCCKGTYAAYGCKGGYGGYGGYGGYWSYGGYGGWGGYWSYGGYGGYGGYWSYGARGGYGGWG 814

Qy         80 GCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGG 139

Db        815 GYWSYGARGGWGGYGGYWSYGGRGGYGGYWSYGGYWSYGGYGAYTTYGAYTAYGARAARA 874

Qy        140 TTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTG 199

Db        875 TGGCWAAYGCYAAYAARGGSGCYATGACYGARAAYGCYGAYGARAAYGCRCTRCARWSTG 934

Qy        200 AAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATA 259

Db        935 AYGCYAARGGYAARYTWGAYWSYGTCGCYACWGAYTAYGGTGCTGCYATCGAYGGYTTYA 994

Qy        260 ATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGA 319

Db        995 TYGGYGAYGTYWSYGGYCTKGCTAAYGGYAAYGGWGCYACYGGWGAYTTYGCWGGYTCKA 1054
```

```
Qy        320 ACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              |  |: |:|   |: |  |:  :| :||  |  |:
Db       1055 ATTCYCARATGGCYCARGTYGGWGAYGGKGAYAAY 1089
```

RESULT 4
US-10-021-698A-1216/c
; Sequence 1216, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4044US1
; CURRENT APPLICATION NUMBER: US/10/021,698A
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/211,749
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 6160
; SOFTWARE: Patentin 2.1
; SEQ ID NO 1216
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-021-698A-1216

Query Match 6.9%; Score 31.4; DB 6; Length 435;
Best Local Similarity 46.2%; Pred. No. 0.51;
Matches 104; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```
Qy         40 GTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAAC 99
              | |  ||   |||||   | ||| |  | |     |  | |  |  || ||||||  | ||
Db        353 GATGGTGATGGTGCTGATGGTGGTGATGATAGTGATGGTGATGATGGTGGTAATGGTGAT 294

Qy        100 GGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCAT 159
              |||   ||  || ||   |     ||   | || |    | | | ||  || |    ||
Db        293 GGCAATGGTGATGGTGGTGATGGTGATGGTAATGATGGTGGCATCAGTAATGGTGGTGAT 234

Qy        160 GAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGC 219
              || |  |    ||       | | ||||  || |   |||       | ||   |
Db        233 GATAGTGATGGTGATGATAATGGTGATGATGGTGATGATGATGGTGATAATAGTGATGAT 174

Qy        220 GGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              ||| ||||| | |  | ||||  ||  |   ||   || || ||
Db        173 GGTAATGGTGATGATGGTGATGATGGTGACCATGTTAATGATGGT 129
```

RESULT 5
US-10-775-972-153/c
; Sequence 153, Application US/10775972
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C21
; CURRENT APPLICATION NUMBER: US/10/775,972
; CURRENT FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 563
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153
; LENGTH: 2007
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-972-153

Query Match 6.9%; Score 31.4; DB 6; Length 2007;
Best Local Similarity 46.5%; Pred. No. 1.1;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```
Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||  |  |  ||| || |  ||| |||       ||  | |  ||   |  |||    |
Db       1437 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1378

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              |      | |  | ||||  |  |  |    |||   ||  || |||  |  ||| |   |
Db       1377 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1318

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              ||||  |    |  | || |  |      |    ||| || ||| |    | |    |||
Db       1317 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1258

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              ||    ||  | |  ||||| | |  | || |||| |
Db       1257 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1221
```

RESULT 6
US-10-775-972-154/c
; Sequence 154, Application US/10775972
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C21
; CURRENT APPLICATION NUMBER: US/10/775,972
; CURRENT FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 563
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 2148
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-972-154

Query Match 6.9%; Score 31.4; DB 6; Length 2148;
Best Local Similarity 46.5%; Pred. No. 1.1;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```
Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||  |  |  ||| || |  ||| |||       ||  | |  ||   |  |||    |
Db       1570 GCTTGTTCATTGAACACCATGTTCCCCAGAAAATTCTCTACAAAGATGTTCATTTTTTTA 1511

Qy        179 AAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCG 238
              |      | |  | ||||  |  |  |    |||   ||  || |||  |  ||| |   |
Db       1510 AGTTAAAAGTACTTAAAAGTTCACTGGGAATTTTGCAATGCAGTTGAGAAAACTGGAGCT 1451

Qy        239 ATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATA 298
              ||||  |    |  | || |  |      |    ||| || ||| |    | |    |||
Db       1450 GCCTAGGGGAATTTCAGCATTAGGGCTTGAAATTACTCACCCAGGAGAAATCCCAGTATA 1391

Qy        299 ATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGA 335
              ||    ||  | |  ||||| | |  | || |||| |
Db       1390 ATCTGCCCTCCCATTAGTGGTAGGAGAGAAGCTCCAA 1354
```

RESULT 7
US-10-021-698A-796/c
; Sequence 796, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON

```
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 796
;   LENGTH: 9278
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-021-698A-796

  Query Match             6.8%;  Score 31.2;  DB 6;  Length 9278;
  Best Local Similarity   47.1%;  Pred. No. 2.6;
  Matches   96;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         85 GGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACC 144
              || | | || ||   | ||| ||  || ||   ||    ||   |||| |    |
Db        347 GGTGATGATGATGGTGACGGTGGTGATGGTGATGGTGATGATGGTGATGATGGTGACGGT 288

Qy        145 CGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACG 204
              ||| || | |  ||||   |  ||||        | || | ||   |||
Db        287 GGTGATGGTGCTGATGATGATGGTGATGCTGATGGTGATGGTGACGGTGATGATGATGGT 228

Qy        205 ACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGT 264
              |  |     |  | | | ||||| | || |  | ||  ||    ||||  ||| || ||
Db        227 GACGGTGATGATGGTGATGATGGTGATGGTGATGCTGATGGTGGTGGTGGTGATGATGGT 168

Qy        265 ACTATTGAACTGACTCAGAATGGT 288
              |  |||     | |  |||||
Db        167 GGTGATGATGATGATGATGATGGT 144


RESULT 8
US-10-767-701-8827
; Sequence 8827, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 8827
;   LENGTH: 1099
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1099)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS59062_1
US-10-767-701-8827

  Query Match             6.8%;  Score 31;  DB 6;  Length 1099;
  Best Local Similarity   50.0%;  Pred. No. 1.1;
  Matches   76;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy         99 CGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCA 158
              ||||||||||    || | |  |  ||   |  ||| | |  ||  | |    || |
Db        277 CGGCGGCGGCCTCGGTGCGGTTCTTGATCCGGGGGAGCGAGCTCACCTGGGTCGCGCCAA 336

Qy        159 TGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAG 218
              || |   | ||  |  | || | ||       |  ||   ||| |  ||  | ||

Db        337 TGCAGCCGGACCCGAGCGGCGACGCCGGCNATGCGAAGGCGAAGCTGCCGCTGCCGCAGC 396

Qy        219 CGGTTATGGTAACGGCGCCGATGTAGGCCAGG 250
              |||| | ||   ||||||||  || ||   ||
Db        397 CGGTGACGGCTCCGGCGCCGTCGTCGGGGCGG 428


RESULT 9
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
;  APPLICANT: Neubauer, Antonie
;  APPLICANT:  Ziegler, Christina
;  TITLE OF INVENTION: gM-Negative EHV-Mutants without Heterologous Elements
;  FILE REFERENCE: 1/1372
;  CURRENT APPLICATION NUMBER: US/10/624,149A
;  CURRENT FILING DATE:  2003-07-21
;  PRIOR APPLICATION NUMBER: 60/403,282
;  PRIOR FILING DATE: 2002-08-14
;  PRIOR APPLICATION NUMBER: DE 10233064
;  PRIOR FILING DATE: 2002-07-19
;  PRIOR APPLICATION NUMBER: DE 10317008
;  PRIOR FILING DATE: 2003-04-11
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
;   LENGTH: 150223
;   TYPE: DNA
;   ORGANISM: Equine herpesvirus 1
;  PUBLICATION INFORMATION:
;   AUTHORS: Telford,E.A.
;   AUTHORS: Watson, M.S.
;   AUTHORS: McBride, K.
;   AUTHORS: Davison, A.J.
;   TITLE: The DNA sequence of equine herpesvirus-1
;   JOURNAL: Virology
;   VOLUME: 189
;   ISSUE: 1
;   PAGES: 304-316
;   DATE: JUL-1992
;   DATABASE ACCESSION NUMBER: NC 001491, NCBI
;   DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

  Query Match             6.8%;  Score 31;  DB 6;  Length 150223;
  Best Local Similarity   51.0%;  Pred. No. 11;
  Matches   73;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        176 TGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCG 235
              ||  ||  |  | || ||  | ||| ||| |   |  ||| |||   | ||   ||| |
Db      95545 TGGTAACGGTGGGCCTTAGTTTTGATACGGCTCATTTCCACAGCATGTTTGCCTACGTCA 95604

Qy        236 CCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
                |  |||| ||  ||||||    | || | | | | ||| |    |  |||||
Db      95605 AGCCCATGAGCCACGGACCGGATATGGTTTCTGTATACCAGTCTCTGGGGCGCGTCAGAG 95664

Qy        296 ATAATGCCACCATCGACCAGTGG 318
              |  |     || ||| | || |
Db      95665 AGCTTATTCACAACGAGCTGTTG 95687


RESULT 10
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
;  APPLICANT: Bayer Pharmaceuticals Corporation
;  APPLICANT:  Eveleigh, Deepa
;  APPLICANT:  Bigwood, Douglas
;  TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
;  FILE REFERENCE: 5152
;  CURRENT APPLICATION NUMBER: US/10/788,792
;  CURRENT FILING DATE:  2004-02-27
;  PRIOR APPLICATION NUMBER: US 60/450,655
```

```
;  PRIOR FILING DATE: 2003-02-28
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
;   LENGTH: 1575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-788-792-27

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 1575;
  Best Local Similarity   55.7%;  Pred. No. 1.5;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
              |||   | | | | || ||   | | || |  || | |||    || |||| ||||||||
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
              |  | |   | |  |||  |  |   |   |||  |  ||||||||
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583


RESULT 11
US-10-767-701-1287/c
; Sequence 1287, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 1287
;   LENGTH: 883
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(883)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS19621_1
US-10-767-701-1287

  Query Match             6.7%;  Score 30.4;  DB 6;  Length 883;
  Best Local Similarity   61.2%;  Pred. No. 1.6;
  Matches   49;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              | ||||| || | || |||||        |||  |||   |  ||| ||  | |||||
Db        493 GGTCTGGATGACCTCATTCCACTGCTCAACGGTAGCGTCGAGGATACCGTTGTCGGCAGT 434

Qy        112 AGTTCCGGCCCGGACTATGA 131
              |||||||   | | ||||
Db        433 CTTTCCGGCGTTGGCGATGA 414


RESULT 12
US-10-767-795-6329/c
; Sequence 6329, Application US/10767795
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Zhou, Yihua
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53534)B
;  CURRENT APPLICATION NUMBER: US/10/767,795
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 6329
;   LENGTH: 989
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C1002_1
US-10-767-795-6329

  Query Match             6.6%;  Score 30.2;  DB 6;  Length 989;
  Best Local Similarity   49.7%;  Pred. No. 1.9;
  Matches   77;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy         24 ATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGG 83
              ||   ||| ||  | |  ||| |  | || ||  |||||| | ||| |          | |
Db        842 ATAAACAGAAACAGCATCTTCAGATAATGATCCTGCTGGCCTTGTGCGGTGCGATGATGA 783

Qy         84 CGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTAC 143
              || |||  |  | | |  ||  | |||     ||||   |  ||   |  |||  |
Db        782 TGGTGGTGGTGGTGATGATGGTATCCATAAGCATGGCCACCAAGATATGAACTTGGAAAG 723

Qy        144 CCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              |||| | |  | |||  |  |||  ||  ||  ||
Db        722 CCGTATGGAAAACCAGCAGCTGGACCAAACATAGC 688


RESULT 13
US-10-767-701-12486/c
; Sequence 12486, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 12486
;   LENGTH: 1203
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS13462_1
US-10-767-701-12486

  Query Match             6.6%;  Score 30;  DB 6;  Length 1203;
  Best Local Similarity   48.3%;  Pred. No. 2.5;
  Matches   84;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy         20 CAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGG 79
              | | |  |||   | ||| || | | | | |||  || | ||  |  ||   |  |   |
Db        271 CCGGAACCGCCTAAGTCGCAGGTGCCGCCGGTGGCCTTGTTGTAGATGTAGTAGGAGTCG 212

Qy         80 GCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGG 139
              ||||  |||   | |   ||||| |||   | ||    |   | | |   |  | || ||
Db        211 CCGGCCCCGGAGGTGAGCGCGGCGACGGTGTTTGTGTTGTAGCAGCCGCCGCCCTGCAGG 152

Qy        140 TTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTA 193
              |   ||||  ||   |  |   ||      |||| | ||  || |  |  ||
Db        151 ATGGACGTGCAGTCGGCGCCGCCCTGCGGACCGCACGGGAAGTCGCTCGCTTTA 98


RESULT 14
US-10-767-701-6233/c
; Sequence 6233, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
```

```
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 6233
;   LENGTH: 1228
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS37536_1
US-10-767-701-6233

  Query Match             6.5%;  Score 29.8;  DB 6;  Length 1228;
  Best Local Similarity   52.9%;  Pred. No. 2.9;
  Matches   64;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy         59 CTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCG 118
              ||||||| |    |  | |||  | |||||||  ||  |   |||||  |||  |    | |
Db        666 CTGGCGTAGCCGAAGTAGGGGATGACGGCGGTGATGTTCTTGGCGGAGGCCCTCCGGCAG 607

Qy        119 GCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGC 178
              ||   ||  |||||  ||  | |       |||  |          | ||||  ||  ||||
Db        606 GCGTCGATCATGATGAGGAGCTCCATGAGGTTCTCGTTGGCTGGAGGGCAGGTGGGCTGC 547

Qy        179 A 179
              |
Db        546 A 546


RESULT 15
US-10-767-701-20611/c
; Sequence 20611, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 20611
;   LENGTH: 656
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 11064914
US-10-767-701-20611

  Query Match             6.5%;  Score 29.6;  DB 6;  Length 656;
  Best Local Similarity   52.4%;  Pred. No. 2.5;
  Matches   65;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy         28 GCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGC 87
              | || | | || |   ||| |  || |  ||  | || || ||||     ||  | || |
Db        126 GTAGAAGTTGTCGAGCCTGCCCTTGTTGAGGTCGTCGGCGATCCATGGGTGGCGGTCGCC 67

Qy         88 GGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGT 147
              |||  | | |      |||||||  |      |||  | ||| | || |  | |   |||
Db         66 GGTGACCTTGTGCACGGCGGCGACGAGCGCGGAGCTGACGAAGAACCCCAGCGACAGCGT 7

Qy        148 GTTG 151
              | ||
Db          6 GCTG 3


Search completed: March 19, 2004, 03:10:15
Job time : 55.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-17
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	232.6	51.0	680	12	BJ618688	BJ618688 BJ618688
c	2	37.4	8.2	484	9	AV430994	AV430994 AV430994
	3	37	8.1	590	14	CF848025	CF848025 psMA003xI
	4	37	8.1	610	14	CF842541	CF842541 psHB020xG
c	5	36.6	8.0	891	29	CNS02I89	AL198594 Tetraodon
	6	36.4	8.0	787	28	BZ568375	BZ568375 pacs2-164
	7	35.2	7.7	417	29	FR0020317	AL013202 F.rubripe
c	8	35.2	7.7	811	28	BZ469000	BZ469000 BOOAO58TF
	9	35.2	7.7	813	28	BZ204853	BZ204853 CH230-390
c	10	35	7.7	490	14	CA588919	CA588919 hab51c12.
	11	34.8	7.6	480	9	AU085223	AU085223 AU085223
c	12	34.6	7.6	406	9	AU241511	AU241511 AU241511
c	13	34.6	7.6	499	12	BJ006097	BJ006097 BJ006097
c	14	34.6	7.6	561	12	BJ001420	BJ001420 BJ001420
c	15	34.6	7.6	640	12	BJ000263	BJ000263 BJ000263
c	16	34.6	7.6	656	12	BJ004146	BJ004146 BJ004146
c	17	34.6	7.6	662	12	BJ010653	BJ010653 BJ010653
c	18	34.6	7.6	662	12	BJ011743	BJ011743 BJ011743
c	19	34.6	7.6	671	12	BJ012818	BJ012818 BJ012818
	20	34.6	7.6	682	14	CF402142	CF402142 RTWW1_20_
c	21	34.6	7.6	691	12	BJ000180	BJ000180 BJ000180
	22	34.6	7.6	702	12	BG441624	BG441624 GA__Ea001
	23	34.6	7.6	937	12	BG444255	BG444255 GA__Ea002
c	24	34.4	7.5	365	9	AA746477	AA746477 nw62e02.s
c	25	34.4	7.5	500	14	CA711477	CA711477 wdk2c.pk0
c	26	34.4	7.5	827	28	BH438448	BH438448 BOGVQ28TR
	27	34.2	7.5	296	9	AV058630	AV058630 AV058630
	28	34.2	7.5	523	12	BJ334624	BJ334624 BJ334624
c	29	34	7.5	588	14	CA731849	CA731849 wlp1c.pk0
c	30	33.8	7.4	522	14	CA707989	CA707989 wdk2c.pk0
	31	33.8	7.4	625	13	BU882267	BU882267 UM74TG08
c	32	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	33	33.6	7.4	584	28	AQ398232	AQ398232 mgxb0001F
	34	33.6	7.4	594	12	BJ335653	BJ335653 BJ335653
	35	33.6	7.4	653	12	BJ336964	BJ336964 BJ336964
c	36	33.6	7.4	712	14	CF437982	CF437982 EST674327
c	37	33.6	7.4	866	29	CNS02C24	AL190597 Tetraodon
	38	33.6	7.4	896	28	CC333807	CC333807 OGMBC13TV
c	39	33.6	7.4	910	14	CA266034	CA266034 SCJLRT307
c	40	33.4	7.3	525	9	AU242732	AU242732 AU242732
c	41	33.4	7.3	596	9	AU177600	AU177600 AU177600
	42	33.4	7.3	649	12	BI261742	BI261742 602954071
	43	33.4	7.3	679	29	CC961526	CC961526 BOIGS10TF
c	44	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
c	45	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 51.0%; Score 232.6; DB 12; Length 680;
Best Local Similarity 69.3%; Pred. No. 4.7e-54;
Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
AV430994/c
LOCUS AV430994 484 bp mRNA linear EST 23-AUG-2000
DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone PL027e10_r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
REFERENCE 1 (bases 1 to 484)
AUTHORS Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and Tabata,S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
PUBMED 10907854
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PL027e10_r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 15;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |   | |  | | ||  ||| |||||   ||| ||||| ||||||  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
              | ||  ||   ||  |||| ||| ||||||      |  ||    |||  |  || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323
```

RESULT 3
CF848025
LOCUS CF848025 590 bp mRNA linear EST 30-OCT-2003
DEFINITION psMA003xI14f USDA-IFAFS:Expression of Phytophthora sojae genes during infection and propagation_sMA Phytophthora sojae cDNA clone sMA003I14 5, mRNA sequence.
ACCESSION CF848025
VERSION CF848025.1 GI:38063679
KEYWORDS EST.
SOURCE Phytophthora sojae
ORGANISM Phytophthora sojae
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.
REFERENCE 1 (bases 1 to 590)
AUTHORS Tyler,B.
TITLE Tyler,B. Not Published
JOURNAL Unpublished (2003)
COMMENT Contact: Tyler B
Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtyler@vt.edu
PCR PRimers
FORWARD: BK reverse primer
BACKWARD: BK reverse primer
Plate: 003 row: I column: 14
Seq primer: BK reverse primer
High quality sequence stop: 590.
FEATURES Location/Qualifiers
source 1. .590
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="sMA003I14"
/tissue_type="mycelium"
/cell_line="P6497"
/dev_stage="mycelium"
/lab_host="synthetic medium"
/clone_lib="USDA-IFAFS:Expression of Phytophthora sojae genes during infection and propagation_sMA"
/note="Vector: pBK-CMV; Site_1: EcoR1; Site_2: Xho1"
ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 590;
Best Local Similarity 51.5%; Pred. No. 21;

```
  Matches   85;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy         26 TCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCG 85
              | | |||  | ||  ||  || | |||  || |  |  ||||      ||   | |||||
Db        224 TTGTAGCGGTAGTCCTTGATGTCCGTGAGCTTGGGGAAGTCGCCGATGAAGTAGCCGGCG 283

Qy         86 GCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCC 145
              || | || |  ||| ||||| | | |    | |       | |||||   | ||    |
Db        284 CCGTTGATGAGCACGACGGCGCCCACAAAGGCCGTGGCCGCGATGATGGCCAGGCGCTCG 343

Qy        146 GTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCC 190
              | |||    |||  |    ||| |||  |  ||   ||||| ||
Db        344 ATCTTGAAGGCCAGCAGGCCGCAGATGAGCCCCAGCACGATGGCC 388


RESULT 4
CF842541
LOCUS       CF842541                610 bp    mRNA    linear   EST 30-OCT-2003
DEFINITION  psHB020xG21f USDA-IFAFS:Expression of Phytophthora sojae genes
            during infection and propagation_sHB Phytophthora sojae cDNA clone
            sHB020G21 5, mRNA sequence.
ACCESSION   CF842541
VERSION     CF842541.1  GI:38058195
KEYWORDS    EST.
SOURCE      Phytophthora sojae
  ORGANISM  Phytophthora sojae
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Tyler,B.
  TITLE     Tyler,B. Not Published
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Tyler B
            Tyler lab
            VBI
            1880 Pratt Dr., Blacksburg, VA 24061, USA
            Tel: 540-231-7318
            Email: bmtyler@vt.edu
            PCR PRimers
            FORWARD: BK reverse primer
            BACKWARD: BK reverse primer
            Plate: 020  row: G  column: 21
            Seq primer: BK reverse primer
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Phytophthora sojae"
                     /mol_type="mRNA"
                     /db_xref="taxon:67593"
                     /clone="sHB020G21"
                     /tissue_type="mycelium"
                     /cell_line="P6497"
                     /dev_stage="48 hr. post infection stage"
                     /lab_host="Soybean plant"
                     /clone_lib="USDA-IFAFS:Expression of Phytophthora sojae
                     genes during infection and propagation_sHB"
                     /note="Vector: pBK-CMV; Site_1: EcoR1; Site_2: Xho1"
ORIGIN

  Query Match             8.1%;  Score 37;  DB 14;  Length 610;
  Best Local Similarity   51.5%;  Pred. No. 22;
  Matches   85;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy         26 TCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCG 85
              | | |||  | ||  ||  || | |||  || |  |  ||||      ||   | |||||
Db        212 TTGTAGCGGTAGTCCTTGATGTCCGTGAGCTTGGGGAAGTCGCCGATGAAGTAGCCGGCG 271

Qy         86 GCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCC 145
              || | || |  ||| ||||| | | |    | |       | |||||   | ||    |
Db        272 CCGTTGATGAGCACGACGGCGCCCACAAAGGCCGTGGCCGCGATGATGGCCAGGCGCTCG 331

Qy        146 GTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCC 190
              | |||    |||  |    ||| |||  |  ||   ||||| ||
Db        332 ATCTTGAAGGCCAGCAGGCCGCAGATGAGCCCCAGCACGATGGCC 376


RESULT 5
CNS02I89/c
LOCUS       CNS02I89                891 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            141G11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL198594
VERSION     AL198594.1  GI:7836745
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 891)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .891
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="141G11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG141AD06SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             8.0%;  Score 36.6;  DB 29;  Length 891;
  Best Local Similarity   51.2%;  Pred. No. 33;
  Matches   66;  Conservative    7;  Mismatches   56;  Indels    0;  Gaps    0;

Qy         51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
              ||  ||| | ||||| |    :  | |||  | |||||||  |        | || ||| :
Db        437 TGTCCTGCCGGGCGTAGGGGAWGCAGGGGATGACGGCGGTKACGCGTGACGAGGAGGCGW 378

Qy        111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
              |  | | |||   |: :|||:||||||| |    | :|||||     :|      |||
Db        377 TCTTGCAGGCGTTGWTYATGWTCAGCTGCTCCWTCAWGTTGTCGTTAWTTTCGCCGCAGC 318

Qy        171 TGCACTGCA 179
              || |||||
Db        317 CGCTCTGCA 309
```

```
RESULT 6
BZ568375
LOCUS       BZ568375                787 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_7503.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_7503, genomic survey sequence.
ACCESSION   BZ568375
VERSION     BZ568375.1  GI:27201285
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .787
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_7503"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 28;  Length 787;
  Best Local Similarity   47.4%;  Pred. No. 36;
  Matches  109;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy         17 TGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAAT 76
              | |  ||  |||  |   ||||  |    | |     || |  || ||  |||  |  |
Db        146 TTGGCGCGCTCGGCGACGTCGTGGAAGGCGTCGAACATCAGCAGGGTGTGCTTCAGCTGT 205

Qy         77 GGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGC 136
              |||||||| | |   ||   ||||| || | |       || || ||      |||
Db        206 TCGGCGGCGGTGTTGGCCAGTTCGGCGTCGTCCAGCAGCTCGACCAGGGTGGCGATGTTG 265

Qy        137 TGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAAT 196
              | |   |||    || | ||||  |  | |       |  |||    ||||    |  |
Db        266 TAGCCGCCCTGCATGGTGCCCAGCAGCTCGGCGGCACGTTGCTTGTCGATAAGGGGGGAG 325

Qy        197 CTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
               ||      || ||  | |  ||||  | |  | ||||    || ||||||
Db        326 GTGGCTTCGCCCTTGACGATGGCGGAAAGGAAACCGGCCTTGACGTAGGC 375


RESULT 7
FR0020317
LOCUS       FR0020317               417 bp    DNA     linear   GSS 10-DEC-1997
DEFINITION  F.rubripes GSS sequence, clone 040D20bD10, genomic survey sequence.
ACCESSION   AL013202
VERSION     AL013202.1  GI:2679570
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
                     /clone="040D20bD10"
                     /clone_lib="cosmid 040D20"
ORIGIN

  Query Match             7.7%;  Score 35.2;  DB 29;  Length 417;
  Best Local Similarity   50.3%;  Pred. No. 59;
  Matches   76;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        170 ATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTA 229
              |||||   ||||  |  ||  |   | ||| |  |||  |   || ||   |||||  || |
Db         16 ATGCAANCCAAAAAGGAGCACCNCANTCCGTCACGCGCNCAACGCACGCCGGTTCAGGGA 75

Qy        230 ACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTT 289
              |    |||  |||    |  |  || |     | |   |  |  |  |||||  |    |
Db         76 ATTCNGCCTGTGTTCACACGCCTGTGTTCTTCATTTGGACTCCAAGGACTCTAATACNTG 135

Qy        290 TCAGAAATAATGCCACCATCGACCAGTGGAA 320
              |||||  ||   ||||  ||||   ||||
Db        136 AGAGAAACGATTATACCACGGACCTNAGGAA 166


RESULT 8
BZ469000/c
LOCUS       BZ469000                811 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BOOAO58TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOAO58,
            genomic survey sequence.
ACCESSION   BZ469000
VERSION     BZ469000.1  GI:26764546
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOOAO58TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOOAO58"
                     /clone_lib="BO_1.6_2_KB_tot"
```

/note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 811;
Best Local Similarity 47.7%; Pred. No. 78;
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```
Qy      146 GTGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGA 205
            || |||  |   | ||    | |  |    ||  | || |     ||||| |||
Db      640 GTATTGAAAATAAGGATTGTGAAGCTTGCATGTTAGGCAAGCATTGTAAAACTGTGTTTC 581

Qy      206 CCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTA 265
            |||||  |    |  | |||||  ||  ||   ||| ||   ||   ||  ||  |
Db      580 CCATTTCATCTACTGTTTATGACAAATGCTTTGATCTAATTCACTCTGATGTTTGGACTG 521

Qy      266 CTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325
            ||  |    || ||  | || |||| | | |   || |||  ||    |   ||  | |
Db      520 CTCCTTGCTTGTCTAGGGATAGTTTTAAATACTTTGTCACATTCATAAATGAAAAATCGA 461

Qy      326 AAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
            || || ||   || ||  |  |||  |   | ||
Db      460 AATACACCTGGATAACACTTATTCAAACCAAAGATA 425
```

RESULT 9
BZ204853
LOCUS BZ204853 813 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-390D21, genomic survey sequence.
ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 813)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-390D21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 390 row: D column: 21
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .813
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"
ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 813;
Best Local Similarity 49.0%; Pred. No. 78;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```
Qy      178 CAAAGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCC 237
            ||| |   | | | | || |   |   ||    ||  | ||  |  |   ||  |  ||
Db      548 CAAGGACCTCCACATCAAACCAGACACACTCAAACTAATAGAAGAAAAACTAGGGAAGCA 607

Qy      238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
              || |   || ||       || |  | |  |||||  |     |||||  ||  |   |
Db      608 TCTGGAACACATGGGCACTGGAAAAAAATTCCTGAACAAAACACCAATGGCTTATGCTCT 667

Qy      298 AATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGC 357
            ||   ||  |||||||| ||| |     | |||||  | |   | ||||  ||| ||     |
Db      668 AAGATCAAGATCGACCAATGGGATCTCATAAACTGCAAAGCTTCTGTAGGCAAAGGACAC 727

Qy      358 GGTAATAACGCC 369
              ||  | | | |
Db      728 TGTGGTTAGGAC 739
```

RESULT 10
CA588919/c
LOCUS CA588919 490 bp mRNA linear EST 19-NOV-2002
DEFINITION hab51c12.y1 Fugu UT7 adult skin Takifugu rubripes cDNA clone IMAGE:6352126 5' similar to SW:CLD4_MOUSE O35054 CLAUDIN-4 ;, mRNA sequence.
ACCESSION CA588919
VERSION CA588919.1 GI:25133497
KEYWORDS EST.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 490)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
Library materials provided by G. Elgar (UK MRC HGMP-RC) Library constructed by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi and S. Watabe (University of Tokyo, Institute of Medical Science and Laboratory of Aquatic Molecular Biology and Biotechnology) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Fugu clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham.
FEATURES Location/Qualifiers
source 1. .490
/organism="Takifugu rubripes"
/mol_type="mRNA"
/db_xref="taxon:31033"
/clone="IMAGE:6352126"
/sex="female"
/tissue_type="skin"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="Fugu UT7 adult skin"

/note="Vector: pBluescript-FL; Site_1: PflM I (CCANNNNNTGG); Site_2: PflM I (CCANNNNNTGG); BamHI-SmaI sites were converted to BamHI-PflMI-SfiI-PflMI sites (SmaI is destroyed). Other part of the vector is untouched. The cDNA is inserted between two PflMI sites in T3 (5') to T7 (3') direction. Library materials provided by G. Elgar (UK MRC HGMP-RC) and constructed and donated by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi and S. Watabe (University of Tokyo, Institute of Medical Science and Laboratory of Aquatic Molecular Biology and Biotechnology)."

ORIGIN

Query Match 7.7%; Score 35; DB 14; Length 490;
Best Local Similarity 51.6%; Pred. No. 72;
Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy        92 ATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTG 151
             ||||| | |||| |||| | ||||| | | |||| ||   |
Db       307 ATCATGAGGGCGCGGGCAGCCTGGAGGTCCCGGGAAAGGGCCAGCAGGGAGTCGTAAACC 248

Qy       152 TTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGACCATTA 211
             || | |   | || |  ||| ||||| || | | | |  | | | | ||| |
Db       247 TTGCACTGCATCTGACCTGTGCTCTGCACCACGCAGTTCATCCAGATCCCTTCCCAAGTG 188

Qy       212 CCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGC 246
                 |||| || ||| |||||| ||||| | ||||
Db       187 ATTTGAGCCGTGATGATAACGGGGCCGACGAAGGC 153
```

RESULT 11
AU085223
LOCUS AU085223 480 bp mRNA linear EST 01-SEP-2000
DEFINITION AU085223 Cryptomeria japonica inner bark Cryptomeria japonica cDNA clone CC2407 5', mRNA sequence.
ACCESSION AU085223
VERSION AU085223.1 GI:9963206
KEYWORDS EST.
SOURCE Cryptomeria japonica (Japanese cedar)
ORGANISM Cryptomeria japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.
REFERENCE 1 (bases 1 to 480)
AUTHORS Ujino-Ihara,T., Yoshimura,K., Ugawa,Y., Yoshimaru,H., Nagasaka,K. and Tsumura,Y.
TITLE Expression analysis of ESTs derived from the inner bark of Cryptomeria japonica
JOURNAL Plant Mol. Biol. 43 (4), 451-457 (2000)
MEDLINE 20505378
PUBMED 11052197
COMMENT Contact: Tokuko Ujino-Ihara
Bio-resources Technology Division
Forestry and Forest Products Research Institute
Matsunosato 1, Kukizaki, Ibaraki 305-8687, Japan
Tel: 81-298-73-3211(ex.444)
Fax: 81-298-73-3795
Email: udino@ffpri.affrc.go.jp;
URL:http://www.ffpri.affrc.go.jp/labs/cjgenome/database/cjdatae.htm
1.
FEATURES Location/Qualifiers
source 1. .480
/organism="Cryptomeria japonica"
/mol_type="mRNA"
/db_xref="taxon:3369"
/clone="CC2407"
/tissue_type="inner bark"
/clone_lib="Cryptomeria japonica inner bark"

ORIGIN

Query Match 7.6%; Score 34.8; DB 9; Length 480;
Best Local Similarity 51.3%; Pred. No. 81;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy       147 TGTTGTTACCCATGAAATGGCACATGCACTGCAAAGCGATGCCCGTAAATCTGAAACGAC 206
             || | || | ||  ||| | ||   |   ||   || ||  |  |||| ||||| |||
Db       219 TGGTTTTGCTCAACAAAGGCCAGGGGTTATGGGTAGAGAATCAAATAAAACTGAAGGGAC 278

Qy       207 CATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTAC 266
             || ||||| ||| |   | | ||      ||| ||  |  |||   ||||  ||
Db       279 TGCGACACAGAGAGGTCACTTTGAGGGGCATTCTGTTGGAAATTCTGCCCGGAATATGAC 338

Qy       267 TATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCA 304
             |  |      | || ||   |  || | | ||| ||
Db       339 TGGTAGCAATTCACAAAAGTATCAAAGCATTCATGGCA 376
```

RESULT 12
AU241511/c
LOCUS AU241511 406 bp mRNA linear EST 15-JAN-2002
DEFINITION AU241511 UV irradiated OLHNI cell line cDNA library (OLc) Oryzias latipes cDNA clone OLc53.01d similar to pir|T16763|hypothetical protein R151.2 - Caenorhabditis elegans, mRNA sequence.
ACCESSION AU241511
VERSION AU241511.1 GI:18154090
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 406)
AUTHORS Naruse,K., Mitani,H. and Tanaka,M.
TITLE Medaka EST Project in University of Tokyo (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from UV irradiated OLHNI cell line cDNA library (OLc) 5' end sequences.
FEATURES Location/Qualifiers
source 1. .406
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="HNI"
/db_xref="taxon:8090"
/clone="OLc53.01d"
/clone_lib="UV irradiated OLHNI cell line cDNA library (OLc)"

ORIGIN

Query Match 7.6%; Score 34.6; DB 9; Length 406;
Best Local Similarity 54.3%; Pred. No. 86;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy        51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
             ||  ||| | |||||       |  | |||  | |||||||| | |       |  | |||| |
Db       293 TGTCCTGCCGGGCGTACGGGAAGCAGGGGATGACGGCGGTGACCCGGGACGACGAGGCGA 234

Qy       111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
             |  | | |||   ||  |||||||||| | |     |  |||||      ||   |  |||
Db       233 TCTTGCAGGCGTTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC 174

Qy       171 TGCACTGCA 179
             || |||||
Db       173 CGCTCTGCA 165
```

RESULT 13
BJ006097/c
LOCUS BJ006097 499 bp mRNA linear EST 05-DEC-2001

```
DEFINITION  BJ006097 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA088D11 5',
            mRNA sequence.
ACCESSION   BJ006097
VERSION     BJ006097.1  GI:17359814
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA088D11"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 499;
  Best Local Similarity   54.3%;  Pred. No. 94;
  Matches   70;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy         51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
              ||  ||| | |||||       |  | |||  | |||||||| | |       |  | ||| |
Db        195 TGTCCTGCCGGGCGTACGGGAAGCAGGGGATGACGGCGGTGACCCGGGACGACGAGGCGA 136

Qy        111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
              |  | | |||    ||  |||||||||| | |     | |||||      ||   |  |||
Db        135 TCTTGCAGGCGTTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC 76

Qy        171 TGCACTGCA 179
                || |||||
Db         75 CGCTCTGCA 67


RESULT 14
BJ001420/c
LOCUS       BJ001420                561 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ001420 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA007C02 5',
            mRNA sequence.
ACCESSION   BJ001420
VERSION     BJ001420.1  GI:17361687
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA007C02"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 561;
  Best Local Similarity   54.3%;  Pred. No. 98;
  Matches   70;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy         51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
              ||  ||| | |||||       |  | |||  | |||||||| | |       |  | ||| |
Db        230 TGTCCTGCCGGGCGTACGGGAAGCAGGGGATGACGGCGGTGACCCGGGACGACGAGGCGA 171

Qy        111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
              |  | | |||    ||  |||||||||| | |     | |||||      ||   |  |||
Db        170 TCTTGCAGGCGTTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC 111

Qy        171 TGCACTGCA 179
                || |||||
Db        110 CGCTCTGCA 102


RESULT 15
BJ000263/c
LOCUS       BJ000263                640 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ000263 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA014A01 5',
            mRNA sequence.
ACCESSION   BJ000263
VERSION     BJ000263.1  GI:17358166
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA014A01"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match             7.6%;  Score 34.6;  DB 12;  Length 640;
```

```
  Best Local Similarity   54.3%;  Pred. No. 1e+02;
  Matches   70;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          51 TGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAA 110
               ||  ||| | |||||       |  | |||  | ||||||| | |       |  | ||| |
Db         387 TGTCCTGCCGGGCGTACGGGAAGCAGGGGATGACGGCGGTGACCCGGGACGACGAGGCGA 328

Qy         111 TAGTTCCGGCCCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 170
               |  | | |||   ||  |||||||||| | |    |  |||||    ||   |  |||
Db         327 TCTTGCAGGCGTTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC 268

Qy         171 TGCACTGCA 179
                 || |||||
Db         267 CGCTCTGCA 259


Search completed: March 18, 2004, 20:11:31
Job time : 2036.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic --nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	412.8	90.5	456	6	I44909	I44909 Sequence 58
	2	412.8	90.5	2067	1	SEU43280	U43280 Salmonella
	3	404.8	88.8	5103	1	STAJ2301	AJ002301 Salmonell
	4	404.8	88.8	22411	1	AE008749	AE008749 Salmonell
	5	403.2	88.4	254050	1	AL627269	AL627269 Salmonell
c	6	403.2	88.4	301983	1	AE016840	AE016840 Salmonell
	7	397.4	87.1	1048	1	STAGFBA	AJ000514 Salmonell
	8	288.6	63.3	361	6	I44908	I44908 Sequence 56
	9	280	61.4	2889	1	CSP515700	AJ515700 Citrobact
	10	238.4	52.3	4680	1	ECCSGABDG	X90754 E.coli csgG
	11	238.4	52.3	10346	1	AE000205	AE000205 Escherich
	12	238.4	52.3	15047	1	D90741	D90741 Escherichia
	13	237.2	52.0	2920	1	CFR515701	AJ515701 Citrobact
	14	236.8	51.9	456	6	AX814811	AX814811 Sequence
	15	236.8	51.9	648	1	ECOCSGAA	L04979 Escherichia
	16	228.6	50.1	1711	1	AF275733	AF275733 Escherich
	17	228.6	50.1	10190	1	AE005315	AE005315 Escherich
	18	228.6	50.1	327773	1	AP002554	AP002554 Escherich
	19	227	49.8	306358	1	AE016759	AE016759 Escherich
	20	195.2	42.8	2883	1	ESA515702	AJ515702 Enterobac
	21	185.8	40.7	230	1	SEU53207	U53207 Salmonella
	22	169.6	37.2	10370	1	AE015131	AE015131 Shigella
	23	169.6	37.2	292504	1	AE016981	AE016981 Shigella
c	24	146.6	32.1	437	1	AF237726	AF237726 Shigella
	25	93.2	20.4	19201	1	D90742	D90742 Escherichia
	26	64.6	14.2	1212	1	ECO131756	AJ131756 Escherich
	27	48.2	10.6	78	6	AX814809	AX814809 Sequence
	28	44.8	9.8	72	6	AX814798	AX814798 Sequence
c	29	42	9.2	2000	6	AX655393	AX655393 Sequence
c	30	36.2	7.9	110243	2	AC146945	AC146945 Oryza sat
c	31	36.2	7.9	148860	8	OSJN00093	AL606648 Oryza sat
	32	35.8	7.9	298900	1	AP005937	AP005937 Bradyrhiz
c	33	35.6	7.8	153232	9	AC024086	AC024086 Homo sapi
c	34	35.6	7.8	185124	9	AC009358	AC009358 Human Chr
c	35	35.6	7.8	186208	9	AC142338	AC142338 Pan trogl
c	36	35.4	7.8	2916	14	FAD459805	AJ459805 Fowl aden
c	37	35.4	7.8	2929	14	FAD554049	AJ554049 Fowl aden
c	38	35.4	7.8	3300	14	FAD431719	AJ431719 Fowl aden
c	39	35.4	7.8	3315	14	FAU26221	U26221 Fowl adenov
c	40	35	7.7	5014	3	PFPFPRNA	X63648 P.falciparu
c	41	35	7.7	12641	1	AE014731	AE014731 Bifidobac
c	42	35	7.7	349980	6	AX492785	AX492785 Sequence
c	43	35	7.7	349980	6	AX553952	AX553952 Sequence
c	44	34.8	7.6	301727	1	AE016960	AE016960 Coxiella
	45	34.8	7.6	347625	1	BX248356	BX248356 Corynebac

ALIGNMENTS

RESULT 1
I44909
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfA gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             90.5%;  Score 412.8;  DB 6;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 6.6e-108;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||    |    ||||||    | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
SEU43280
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match             90.5%;  Score 412.8;  DB 1;  Length 2067;
  Best Local Similarity   94.1%;  Pred. No. 7.2e-108;
```

```
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1193 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 1252

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1253 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 1312

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1313 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 1372

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db       1373 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 1432

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 1492

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1493 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 1552

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1612

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1613 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 1648


RESULT 3
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
```

```
gene            3344. .3376
                /gene="csgBA"
-35_signal      3344. .3349
                /gene="csgBA"
-10_signal      3371. .3376
                /gene="csgBA"
gene            3469. .3924
                /gene="csgB"
CDS             3469. .3924
                /gene="csgB"
                /function="nucleator"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgB protein"
                /protein_id="CAA05316.1"
                /db_xref="GI:2739238"
                /db_xref="GOA:P55226"
                /db_xref="SWISS-PROT:P55226"
                /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
sig_peptide     3469. .3531
                /gene="csgB"
mat_peptide     3532. .3921
                /gene="csgB"
                /product="CsgB protein"
                /function="nucleator"
gene            3966. .4421
                /gene="csgA"
CDS             3966. .4421
                /gene="csgA"
                /function="fiber subunit"
                /codon_start=1
                /evidence=experimental
                /transl_table=11
                /product="CsgA protein"
                /protein_id="CAA05317.1"
                /db_xref="GI:2739239"
                /db_xref="GOA:P55225"
                /db_xref="SWISS-PROT:P55225"
                /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
sig_peptide     3966. .4025
                /gene="csgA"
mat_peptide     4026. .4418
                /gene="csgA"
                /product="CsgA protein"
                /function="fiber subunit"
gene            4483. .4809
                /gene="csgC"
CDS             4483. .4809
                /gene="csgC"
                /codon_start=1
                /transl_table=11
                /product="CsgC protein"
                /protein_id="CAA05318.1"
                /db_xref="GI:2739240"
                /db_xref="SWISS-PROT:P55227"
                /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                WPPSAQ"
sig_peptide     4483. .4533
                /gene="csgC"
mat_peptide     4534. .4806
                /gene="csgC"
                /product="CsgC"
gene            4940. .5103
                /gene="ORF103"
CDS             4940. .>5103
                /gene="ORF103"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA05319.1"
                /db_xref="GI:2739241"
                /db_xref="SWISS-PROT:O54295"
                /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                CDISTHAENI"
ORIGIN

  Query Match             88.8%;  Score 404.8;  DB 1;  Length 5103;
  Best Local Similarity   93.0%;  Pred. No. 1.6e-105;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  |||||| | | | || ||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 4
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
```

```
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
                        /gene="STM1127"
     CDS                complement(1414. .2295)
                        /gene="STM1127"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                        identity in aa 15 - 264"
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="AAL20058.1"
                        /db_xref="GI:16419643"
                        /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                        SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                        SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                        MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                        TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                        QARESKLRTMNALDMTK"
     RBS                complement(2298. .2303)
                        /gene="STM1127"
                        /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene               complement(2580. .4088)
                        /gene="STM1128"
     CDS                complement(2580. .4076)
                        /gene="STM1128"
                        /note="similar to E. coli putative cotransporter
                        (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                        identity in aa 7 - 478"
                        /codon_start=1
                        /transl_table=11
                        /product="putative sodium/glucose cotransporter"
                        /protein_id="AAL20059.1"
                        /db_xref="GI:16419644"
                        /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                        VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                        LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                        LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                        DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                        NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                        AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                        VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                        RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS                complement(4083. .4088)
                        /gene="STM1128"
                        /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene               complement(4413. .5099)
                        /gene="STM1129"
     CDS                complement(4413. .5093)
                        /gene="STM1129"
                        /note="similar to E. coli putative enzyme (AAC76255.1);
                        Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                        226"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20060.1"
                        /db_xref="GI:16419645"
                        /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                        AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                        AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                        ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                        WFSHAVKR"
     RBS                complement(5094. .5099)
                        /gene="STM1129"
                        /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene               5599. .6759
                        /gene="STM1130"
     CDS                5599. .6759
                        /gene="STM1130"
                        /note="similar to E. coli orf, hypothetical protein
                        (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                        identity in aa 40 - 403"
                        /codon_start=1
                        /transl_table=11
                        /product="putative inner membrane protein"
                        /protein_id="AAL20061.1"
                        /db_xref="GI:16419646"
                        /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                        AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                        GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                        EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                        PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                        QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                        WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene               6793. .7497
                        /gene="STM1131"
     RBS                6793. .6798
                        /gene="STM1131"
                        /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
```

```
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
                     (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                     identity in aa 25 - 418"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sugar transport protein"

  Query Match             88.8%;  Score 404.8;  DB 1;  Length 22411;
  Best Local Similarity   93.0%;  Pred. No. 1.7e-105;
  Matches  424;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  |||||| | |  | |  || ||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 5
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
```

```
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap"

  Query Match             88.4%;  Score 403.2;  DB 1;  Length 254050;
  Best Local Similarity   92.8%;  Pred. No. 5.7e-105;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |     | |   || ||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205
```

```
Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361
```

```
RESULT 6
AE016840/c
LOCUS       AE016840              301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
```

```
                RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene       complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
     CDS        complement(5597. .6295)
                /gene="flgH"
                /locus_tag="t1740"
                /note="corresponds to STY1219 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar L-ring protein precursor"
                /protein_id="AAO69364.1"
                /db_xref="GI:29137803"
                /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                MGWLQRFFLNLSPM"
     gene       complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
     CDS        complement(6350. .7132)
                /gene="flgG"
                /locus_tag="t1741"
                /note="distal rod protein; corresponds to STY1218 from
                Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar basal-body rod protein FlgG"
                /protein_id="AAO69365.1"
                /db_xref="GI:29137804"
                /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                QQGQAAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene       complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
     CDS        complement(7146. .7901)
                /gene="flgF"
                /locus_tag="t1742"
                /note="corresponds to STY1217 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative flagellar basal-body rod protein FlgF"
                /protein_id="AAO69366.1"
                /db_xref="GI:29137805"
                /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene       complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
     CDS        complement(7922. .9133)
                /gene="flgE"
                /locus_tag="t1743"
                /note="corresponds to STY1216 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="flagellar hook protein FlgE"
                /protein_id="AAO69367.1"
                /db_xref="GI:29137806"
                /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

Query Match             88.4%;  Score 403.2;  DB 1;  Length 301983;
Best Local Similarity   92.8%;  Pred. No. 5.8e-105;
  Matches  423;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37310 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 37251

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37250 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 37191

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37190 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 37131

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||    |    ||||||      | |     | |    ||  ||
Db      37130 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 37071

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db      37070 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 37011

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      37010 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 36951

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36950 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 36891

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      36890 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 36855


RESULT 7
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
```

```
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             87.1%;  Score 397.4;  DB 1;  Length 1048;
  Best Local Similarity   92.1%;  Pred. No. 1.9e-103;
  Matches  419;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        593 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTGTGGCT 652

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 712

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 772

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||    |    ||||||     | |     | |     ||  ||
Db        773 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 832

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db        833 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 892

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        893 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 952

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 1012

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTA 455
              |||||||||||||||  |||||||||||||| | .
Db       1013 GGTTTTGGCAACAACGCACCGGCTAACCAGTATAA 1047


RESULT 8
I44908
LOCUS       I44908                   361 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 56 from patent US 5635617.
ACCESSION   I44908
VERSION     I44908.1  GI:2469621
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
  TITLE     Methods and compositions comprising the agfA gene for detection of
            Salmonella
  JOURNAL   Patent: US 5635617-A 56 03-JUN-1997;

FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             63.3%;  Score 288.6;  DB 6;  Length 361;
  Best Local Similarity   91.3%;  Pred. No. 4.5e-72;
  Matches  306;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              |||||||||||||| ||| |    |    ||||||     | |     | |     ||  ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 9
CSP515700
LOCUS       CSP515700                2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

  Query Match           61.4%;  Score 280;  DB 1;  Length 2889;
  Best Local Similarity  77.4%;  Pred. No. 1.6e-69;
  Matches  353;  Conservative    0;  Mismatches  100;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||| |||||||||||| |||||||||| ||||||||||||||||||||||
Db       2119 ATGAAACTTTTACAAGTGGCAGCATTTGCAGCAATCGTGGTTTCTGGCAGTGCTCTGGCT 2178

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||    ||||| |||||||||||||||||   |     ||||||||| |  || ||||||
Db       2179 GGTTCTGTTCCGCAATGGGGCGGCGGCGG---TGGCGGCGGCGGCGGGAGCAGCTCCGGC 2235

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||| || ||  |||||||||||||||  ||   |  |||||| ||||||| ||||||
Db       2236 CCGGAATCGACCCTGAGCATTTATCAGTCAGGAGTCAATAACGCCGCGCTTGCGCTGCAA 2295

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||| || ||||||| ||||  |  ||  |  ||     | |   ||| |
Db       2296 AGCGACGCTCGTAAATCTGATACGACCATTCATCAGAATGGCTTTGGTAACGGCGCAGAC 2355

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  ||||||||  | ||||| || || || || |||||||| || || |||| ||| ||
Db       2356 GTGGGCCAGGGCTCAGATAACAGCACCATCGATCTGACTCAAAACGGCTTCAAAAACAAC 2415

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||| |||||||||  ||||| || || |||||||  |||| || || ||
Db       2416 GCCACCATCGATCAGTGGAACGGCAAAAATTCGGACATTACTGTGAGCCAGTATGGTGGA 2475

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||| ||||| || ||||| || || |||||||||||  |||||| |||||||
Db       2476 CATAACGCCGCACTGGTGAACCAGACTGCGTCCGATTCCAGCGTTCTGGTGCATCAGGTT 2535

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||| |||||||||| |||
Db       2536 GGTTTTGGCAACAACGCCACCGCTAACCAGTACTAA 2571
```

```
RESULT 10
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
```

```
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match          52.3%;  Score 238.4;  DB 1;  Length 4680;
  Best Local Similarity  70.2%;  Pred. No. 1.6e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | ||||||||||| | || || || ||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || | .|||||||||||||| ||    || ||||| |||| |  ||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  |||||||||| | |||   |    ||||||    ||     |  | |    ||| ||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                 ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 11
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
```

```
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 (e-mail: mark@amber.gatech.edu). Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
```

```
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

  Query Match           52.3%;  Score 238.4;  DB 1;  Length 10346;
  Best Local Similarity  70.2%;  Pred. No. 1.7e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || |||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||    | ||| |||||| |||||||||   || |||| |||| ||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  |||||||||| | |||   |    ||||||    ||    |  | |    ||| ||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                |||| ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || |||||| |||||||| || ||| | ||||| |||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448


RESULT 12
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                    Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES                    Location/Qualifiers
     source                 1. .15047
                            /organism="Escherichia coli K12"
                            /mol_type="genomic DNA"
                            /strain="K12"
                            /db_xref="taxon:83333"
                            /map="23.7-24.0 min"
                            /clone="Kohara clone #231"
                            /note="Nucleotide position 1097101-1112147 from the
```

```
                initiation site of ThrA (0 min.).~This clone is from
                Kohara lambda miniset library."
     CDS        <1. .276
                /note="unnamed protein product; ORF_ID:o230#5"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35822.1"
                /db_xref="GI:4062605"
                /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
                WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     CDS        1152. .2090
                /note="ORF_ID:o231#1
                similar to SwissProt Accession Number P45637"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical 33.0 kd protein in proB-proA
                intergenic region."
                /protein_id="BAA35823.1"
                /db_xref="GI:4062606"
                /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
                WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
                AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
                PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
                GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
                AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     CDS        2145. .2882
                /note="unnamed protein product; ORF_ID:o231#2"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35824.1"
                /db_xref="GI:4062607"
                /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
                ILNVSPRRLLNFLESRGMAPIAEFADL"
     CDS        2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
     CDS        3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
     gene       complement(4117. .4950)
                /gene="csgG"
     CDS        complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
     gene       complement(4977. .5393)
                /gene="csgF"
     CDS        complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene       complement(5418. .5807)
                /gene="csgE"
     CDS        complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
     gene       complement(5812. .6462)
                /gene="csgD"
     CDS        complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene       7217. .7672
                /gene="csgB"
     CDS        7217. .7672
```

```
  Query Match             52.3%;  Score 238.4;  DB 1;  Length 15047;
  Best Local Similarity   70.2%;  Pred. No. 1.7e-57;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || |||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              | |||||||||| | |||  |    ||||||    ||     |  | |    ||| ||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952
```

```
Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168
```

RESULT 13
CFR515701
LOCUS CFR515701 2920 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter freundii csgB gene, csgA gene and csgD gene.
ACCESSION AJ515701
VERSION AJ515701.1 GI:31790495
KEYWORDS csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation component of curlin monomers; regulatory protein.
SOURCE Citrobacter freundii
ORGANISM Citrobacter freundii
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.
REFERENCE 1
AUTHORS Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
JOURNAL Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE 2 (bases 1 to 2920)
AUTHORS Romling,U.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES Location/Qualifiers
source 1. .2920
/organism="Citrobacter freundii"
/mol_type="genomic DNA"
/isolate="Feç4"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:546"
/country="Germany"
gene complement(211. .861)
/gene="csgD"
CDS complement(211. .861)
/gene="csgD"
/function="necessary for transcription of the csgBA operon"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56673.1"
/db_xref="GI:31790496"
/translation="MFNEVHSIHGHTLLLITKPSLQATALLQHLKQSLALTGKLHNIQ
RSLDDISSSCIVLLDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYRDIENWP
HINGVFYATEDEQRVVSGLQGVLRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene 1624. .2082
/gene="csgB"
CDS 1624. .2082
/gene="csgB"
/function="nucleator"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56674.1"
/db_xref="GI:31790497"
/translation="MKNKLLFMMFTMLGVPGIAAATTSYDLANSEYNFAVNELSKSSF
NQAAIIGQVGTANSANTRQGGSKLLSVISQEGSGNRAKTDQTGSYNFAYIDQTGSSND
ASIKQGSYGNTAVIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR"
gene 2123. .2572
/gene="csgA"
CDS 2123. .2572
/gene="csgA"
/function="curli subunit"
/codon_start=1
/transl_table=11
/product="curlin-csgA protein"
/protein_id="CAD56675.1"
/db_xref="GI:31790498"
/translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGNHHGGGSNYGPDSSLS
IYQYGSNNSANALQSDARKSDVTITQHGRGNGAVVGQGADDSTISLKQTGFQNSATID
QWNAKNADISVTQFGGRNGALVNQTASDSNVLIQQVGFGNNATANQH"
ORIGIN

```
Query Match            52.0%;  Score 237.2;  DB 1;  Length 2920;
Best Local Similarity  71.7%;  Pred. No. 3.5e-57;
Matches  327;  Conservative    0;  Mismatches  123;  Indels    6;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2123 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 2182

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| ||||||||||||      |||||| | || || |||| || ||  |||
Db       2183 GGTGTTGTTCCGCAATGGGGCGGC------AATCATCATGGTGGTGGCAGTAATTATGGC 2236

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              || |||||  |  |||| || || || || || ||    ||| ||||   ||||||||||
Db       2237 CCAGACTCTTCACTGAGTATCTACCAATATGGGTCAAACAACTCTGCAAATGCTCTGCAA 2296

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              || ||||| |||||||  |||       | || |     |          |    ||   |
Db       2297 AGTGATGCGCGTAAATCAGATGTCACCATCACACAACACGGGCGTGGCAACGGAGCTGTT 2356

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  |||||||||| ||| | ||||| |||   ||||  ||||  || |||   || | |
Db       2357 GTTGGCCAGGGTGCTGATGACAGTACCATTAGCCTGAAACAGACCGGCTTCCAGAACAGT 2416

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||| |||| |||||||| || |||||  | |||||||  ||    ||| | ||| ||
Db       2417 GCCACTATCGATCAGTGGAATGCAAAAAATGCTGATATTAGCGTGACCCAGTTCGGTGGC 2476

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||  ||| |||||||||||||| || ||||| ||||  ||  || | |  ||||||
Db       2477 CGCAACGGTGCGTTGGTTAATCAGACAGCCTCTGACTCCAATGTGCTGATTCAGCAGGTT 2536

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||||||||  | |||
Db       2537 GGCTTTGGTAACAACGCTACAGCTAACCAACACTAA 2572
```

RESULT 14
AX814811
LOCUS AX814811 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
TITLE Peptides
JOURNAL Patent: WO 03064446-A 15 07-AUG-2003;
Hansa Medical Research Aktiebolag (SE)
FEATURES Location/Qualifiers
source 1. .456

```
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match             51.9%;  Score 236.8;  DB 6;  Length 456;
  Best Local Similarity   70.0%;  Pred. No. 4.1e-57;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              | ||||||||||| | |||    |     ||||||     ||     | | |     ||| ||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||   || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 15
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.

FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN

  Query Match             51.9%;  Score 236.8;  DB 1;  Length 648;
  Best Local Similarity   70.0%;  Pred. No. 4.1e-57;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              | ||||||||||| | |||    |     ||||||     ||     | | |     ||| ||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||   || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


Search completed: March 18, 2004, 15:02:54
Job time : 2312.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64626	Aac64626 AgfA::PT3
	2	412.8	90.5	456	2	AAQ87467	Aaq87467 AgfA sequ
	3	412.8	90.5	456	2	AAT74142	Aat74142 Salmonell
	4	412.8	90.5	456	3	AAC64617	Aac64617 Salmonell
	5	385.6	84.6	456	3	AAC64629	Aac64629 AgfA::PT3
	6	363.2	79.6	456	3	AAC64625	Aac64625 AgfA::PT3
	7	361.6	79.3	456	3	AAC64628	Aac64628 AgfA::PT3
	8	361.6	79.3	456	3	AAC64622	Aac64622 AgfA::PT3
	9	361.6	79.3	456	3	AAC64623	Aac64623 AgfA::PT3
	10	360	78.9	456	3	AAC64630	Aac64630 AgfA::PT3
	11	360	78.9	456	3	AAC64624	Aac64624 AgfA::PT3
	12	358.4	78.6	456	3	AAC64627	Aac64627 AgfA::PT3
	13	355.2	77.9	456	3	AAC64631	Aac64631 AgfA::PT3
	14	288.6	63.3	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	288.6	63.3	361	2	AAT74141	Aat74141 Salmonell
	16	236.8	51.9	456	3	AAC64619	Aac64619 Escherich
	17	236.8	51.9	456	9	ACF36153	Acf36153 E. coli C
	18	224.4	49.2	646	2	AAQ62647	Aaq62647 Fibronect
	19	162.4	35.6	369	2	AAQ62646	Aaq62646 FNB curli
	20	51.2	11.2	100	7	ACD68807	Acd68807 E. coli K
c	21	50	11.0	78	3	AAC64609	Aac64609 AgfA (SEF
	22	50	11.0	78	3	AAC64610	Aac64610 AgfA (SEF
	23	48.2	10.6	78	9	ACF36151	Acf36151 E. coli C
	24	48	10.5	48	3	AAC64621	Aac64621 Leishmani
	25	45	9.9	78	3	AAC64606	Aac64606 SefA (SEF
	26	44.8	9.8	72	9	ACF36148	Acf36148 E. coli C
c	27	43.4	9.5	78	3	AAC64605	Aac64605 SefA (SEF
	28	43.2	9.5	48	3	AAC64616	Aac64616 S. enteri
	29	42.4	9.3	100	7	ACD68808	Acd68808 E. coli K
c	30	42	9.2	2000	7	ADA71938	Ada71938 Rice gene
c	31	35	7.7	349980	6	ABQ81847	Abq81847 Bifidobac
c	32	34	7.5	1344	2	AAX09007	Aax09007 Hypersens
c	33	34	7.5	1344	3	AAD00669	Aad00669 Erwinia a
c	34	34	7.5	1344	3	AAA14939	Aaa14939 DNA encod
c	35	34	7.5	1344	6	AAL41133	Aal41133 Hypersens
c	36	34	7.5	1344	6	AAD29125	Aad29125 Erwinia a
c	37	34	7.5	1344	6	AAD27016	Aad27016 E. amylov
c	38	34	7.5	1344	6	ABL51711	Abl51711 Erwinia a
c	39	33	7.2	110000	6	ABA90521_17	Continuation (18 o
	40	32.8	7.2	3562	4	ABL06749	Abl06749 Drosophil
	41	32.8	7.2	24081	7	AAD54223	Aad54223 Streptomy
	42	32.8	7.2	52101	7	AAD54217	Aad54217 Streptomy
	43	32.6	7.1	100	7	ACD68809	Acd68809 E. coli K
	44	32.6	7.1	2553	3	AAA09500	Aaa09500 V. furnis
c	45	32.4	7.1	901	4	ABL09481	Abl09481 Drosophil

ALIGNMENTS

RESULT 1
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

```
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-139;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag=  a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;    94WO-IB000207.
XX
PR   26-APR-1993;    93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 412.8;  DB 2;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 3.6e-125;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||||| |||   |   ||||||     | |     | |    || ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
AAT74142
```

```
ID   AAT74142 standard; DNA; 456 BP.
XX
AC   AAT74142;
XX
DT   25-MAR-2003  (revised)
DT   29-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b agfA gene.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag= a
FT                   /label= agfA_gene_fragment
FT                   /transl_except= (pos:367. .369,aa:Pro)
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23570.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 19-112; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene derived from
CC   Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC   diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 412.8;  DB 2;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 3.6e-125;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 3; Length 456;
Best Local Similarity 94.1%; Pred. No. 3.6e-125;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  |  ||||||  | |  | |  ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
AAC64629
ID AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 84.6%; Score 385.6; DB 3; Length 456;
Best Local Similarity 90.4%; Pred. No. 3.1e-116;
Matches 412; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
                | | |||||     |   ||   |  | | ||  | |     | |   ||  ||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 6
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36349.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

  Query Match           79.6%;  Score 363.2;  DB 3;  Length 456;
  Best Local Similarity  87.3%;  Pred. No. 7.3e-109;
  Matches  398;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||       |   ||   |     | | ||  | |    || | ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||      | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

RESULT 7
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

```
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match             79.3%;  Score 361.6;  DB 3;  Length 456;
  Best Local Similarity   87.1%;  Pred. No. 2.5e-108;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |   | |   ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||  |||      ||   ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              || |    || |   |  ||||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64622
ID   AAC64622 standard; DNA; 456 BP.
XX
AC   AAC64622;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36346.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 135; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

  Query Match             79.3%;  Score 361.6;  DB 3;  Length 456;
  Best Local Similarity   87.1%;  Pred. No. 2.5e-108;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||    |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||| |          | | || | ||   |    | | |  |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 9
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 79.3%; Score 361.6; DB 3; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.5e-108;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||    |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                      ||     | ||||  ||         |  || | ||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 78.9%; Score 360; DB 3; Length 456;
Best Local Similarity 86.8%; Pred. No. 8.3e-108;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||    |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                |  |||   |      |  |  ||   |    |      |||  ||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 11
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36348.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

  Query Match            78.9%;  Score 360;  DB 3;  Length 456;
  Best Local Similarity  86.8%;  Pred. No. 8.3e-108;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||   ||    | | | |   | |     |   ||   ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||| |||   |   ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match            78.6%;  Score 358.4;  DB 3;  Length 456;
  Best Local Similarity  86.6%;  Pred. No. 2.8e-107;
  Matches  395;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||| |||   |   ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||  |  |       | ||  |    ||  |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 13
AAC64631
ID   AAC64631 standard; DNA; 456 BP.
XX
AC   AAC64631;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36355.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 355.2;  DB 3;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 3.1e-106;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||| |||   |   ||||||   | |     | |   || ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||   || |||  |   |                |   | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
```

```
PA  (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA  (KING/) KING J.
XX
PI  Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR  WPI; 1994-358275/44.
DR  P-PSDB; AAR62761.
XX
PT  Eliciting an immune response to Salmonella - using attenuated Salmonella
PT  strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS  Disclosure; Fig 7A; 95pp; English.
XX
CC  The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC  agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC  genetic immunization and vaccine compositions, respectively, to elicit an
CC  immune response to Salmonella in animals (e.g. food producing animals)
CC  and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC  AUG-2003 to correct OS field.)
XX
SQ  Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match            63.3%;  Score 288.6;  DB 2;  Length 361;
  Best Local Similarity  91.3%;  Pred. No. 2.3e-84;
  Matches  306;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              |||||||||||| ||| |  |  ||||||  | |  | |  ||  ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 15
AAT74141
ID  AAT74141 standard; DNA; 361 BP.
XX
AC  AAT74141;
XX
DT  25-MAR-2003  (revised)
DT  30-SEP-1997  (first entry)
XX
DE  Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW  Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS  Salmonella enteritidis.
XX
FH  Key             Location/Qualifiers
FT  CDS             1. .360
FT                  /*tag=  a
FT                  /label= agfA_gene_fragment
FT  primer_bind     16. .60
FT                  /*tag=  b
FT                  /label= Primer_TAF5
FT  primer_bind     52. .69
FT                  /*tag=  c
FT                  /label= Primer_TAF3
FT  primer_bind     complement(103. .128)
FT                  /*tag=  d
FT                  /label= Primer_TAF6
FT  primer_bind     complement(294. .312)
FT                  /*tag=  e
FT                  /label= Primer_TAF4
XX
PN  US5635617-A.
XX
PD  03-JUN-1997.
XX
PF  26-APR-1994;   94US-00233788.
XX
PR  26-APR-1993;   93US-00054452.
XX
PA  (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI  Collinson SK,  Kay WW,  Doran JL;
XX
DR  WPI; 1997-309886/28.
DR  P-PSDB; AAW23569.
XX
PT  Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT  enteropathogenic bacteria of the Enterobacteria family.
XX
PS  Claim 1; Col 107-110; 85pp; English.
XX
CC  The present sequence represents an isolated agfA gene fragment derived
CC  from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC  acid can be used to provide diagnostic assays for Salmonella and/or
CC  enteropathogenic bacteria of the family Enterobacteria. It can also be
CC  used to provide proteins and antibodies which can be used for assays. The
CC  nucleic acid sequence can be used to provide probes or primers which can
CC  specifically hybridise to nucleic acid molecules from greater than 99% of
CC  Salmonella strains that are pathogenic to warm-blooded animals relative
CC  to nucleic acid molecules from virtually all other microbial organisms.
CC  (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ  Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match            63.3%;  Score 288.6;  DB 2;  Length 361;
  Best Local Similarity  91.3%;  Pred. No. 2.3e-84;
  Matches  306;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              |||||||||||| ||| |  |  ||||||  | |  | |  ||  ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


Search completed: March 18, 2004, 09:12:00
```

Job time : 289.405 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	412.8	90.5	456	1	US-08-233-788A-58	Sequence 58, Appl
	2	288.6	63.3	361	1	US-08-233-788A-56	Sequence 56, Appl
c	3	34	7.5	1344	3	US-09-120-927-1	Sequence 1, Appli
c	4	34	7.5	1344	4	US-09-431-614-5	Sequence 5, Appli
	5	32.8	7.2	1218	4	US-09-489-039A-403	Sequence 403, App
c	6	32.2	7.1	399	4	US-09-621-976-8976	Sequence 8976, Ap
	7	32	7.0	1152	4	US-09-489-039A-2755	Sequence 2755, Ap
c	8	31.4	6.9	534	4	US-09-489-039A-2195	Sequence 2195, Ap
	9	31.4	6.9	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
c	10	31.2	6.8	549	4	US-09-252-991A-14225	Sequence 14225, A
c	11	31.2	6.8	696	4	US-09-252-991A-14137	Sequence 14137, A
c	12	31	6.8	2052	1	US-08-443-104-5	Sequence 5, Appli
c	13	31	6.8	2052	1	US-08-238-130-6	Sequence 6, Appli
c	14	31	6.8	2052	1	US-08-442-859-5	Sequence 5, Appli
c	15	31	6.8	2052	2	US-08-398-489-5	Sequence 5, Appli
c	16	31	6.8	2052	2	US-08-894-772-1	Sequence 1, Appli
c	17	31	6.8	2052	2	US-09-207-844-1	Sequence 1, Appli
c	18	31	6.8	2052	4	US-09-252-509-1	Sequence 1, Appli
c	19	31	6.8	2052	5	PCT-US95-05534-5	Sequence 5, Appli
	20	30.8	6.8	1662	1	US-08-565-386-2	Sequence 2, Appli
	21	30.8	6.8	2547	3	US-08-508-761B-1	Sequence 1, Appli
	22	30.8	6.8	4529	1	US-08-565-386-1	Sequence 1, Appli
	23	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	24	30.6	6.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	25	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	26	30	6.6	2436	4	US-09-540-236-492	Sequence 492, App
c	27	30	6.6	3895	4	US-08-961-527-201	Sequence 201, App
c	28	30	6.6	65792	4	US-09-596-002-31	Sequence 31, Appl
	29	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
c	30	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
	31	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
	32	29.8	6.5	51259	3	US-08-781-891-209	Sequence 209, App
	33	29.8	6.5	51259	4	US-09-618-166-209	Sequence 209, App
c	34	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
	35	29.4	6.4	912	4	US-09-543-681A-2040	Sequence 2040, Ap
c	36	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
c	37	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
c	38	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
	39	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
	40	29.2	6.4	708	4	US-09-543-681A-2551	Sequence 2551, Ap
	41	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	42	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	43	29	6.4	939	4	US-09-976-594-108	Sequence 108, App
c	44	29	6.4	1097	4	US-09-988-784-7	Sequence 7, Appli
c	45	29	6.4	98844	4	US-09-791-211-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 90.5%; Score 412.8; DB 1; Length 456;
Best Local Similarity 94.1%; Pred. No. 1.7e-137;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| ||| |  |||||| | |  | |  || ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

```
  Query Match           63.3%;  Score 288.6;  DB 1;  Length 361;
  Best Local Similarity 91.3%;  Pred. No. 3.8e-93;
  Matches  306;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              |||||||||||||| ||| |  |||||| | |  | |  || ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335
```

RESULT 3
US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA AND ITS USE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,927
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,108
; FILING DATE: 06-AUG-1977
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1581
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304

```
;       TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1344 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-120-927-1

  Query Match             7.5%;  Score 34;  DB 3;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.09;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT: Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION: RESISTANCE
;  FILE REFERENCE: 21829/41 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/09/431,614
;  CURRENT FILING DATE: 1999-11-02
;  EARLIER APPLICATION NUMBER: 60/107,243
;  EARLIER FILING DATE: 1998-11-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-431-614-5

  Query Match             7.5%;  Score 34;  DB 4;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 0.09;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 5
US-09-489-039A-403
; Sequence 403, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 403
;   LENGTH: 1218
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-403

  Query Match             7.2%;  Score 32.8;  DB 4;  Length 1218;
  Best Local Similarity   58.0%;  Pred. No. 0.23;
  Matches   58;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        174 TCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAAT 233
              ||  || |||| ||  | | ||||||| |||||   ||   |||| |||      || ||
Db        614 TCGTCACAGCGCTGATCCTCAATATGTTCAGCACCTTTATCGTGCTGTCGGTAATTAACC 673

Qy        234 GGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAA 273
              |      || || ||| | || |  ||||  |||   |||
Db        674 CGACGCGGCCGGGCAGCGAGCAGGAAATAAAACTCGAGAA 713


RESULT 6
US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 8976
;   LENGTH: 399
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-8976

  Query Match             7.1%;  Score 32.2;  DB 4;  Length 399;
  Best Local Similarity   10.0%;  Pred. No. 0.19;
  Matches   34;  Conservative  151;  Mismatches  154;  Indels    0;  Gaps    0;

Qy         88 GGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAG 147
              | : ::::    |:::::: :|:  : :: ::   ::::::   |::| :      : :|
Db        340 GWKYWYYWMTSRGSYRRYKTSAMMGRAKMKKRKKTKKMYWMKGGGKKGGSTYMAMRSRRG 281

Qy        148 TACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTG 207
                 :  :   ::: ::::| |:::::  ::: ::::    :  :: :  ::|: ::|: :
Db        280 STGRWSYRRAMWRGSKSWGGGSYYRMAGYRSSRWRSWYSAMWRKKKMTCWKGRSSWGSRS 221

Qy        208 GTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACT 267
              :  : :: : | :  : :: :::  :  |:  : |::|: ::| :|:  : :  |
Db        220 TGYYAWMYKKSWCTSRKWMYYKKRRKKWRRKCTSTKRTCYRGSTYKCWKAYYTKKRRKWT 161

Qy        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327
              ::|   ::: : : : :  :| :::  ::::       |  |  |   :    : :  ::
Db        160 RWTYYYYKSYMSMKKTWRMKTAYYWTKRWKMTRTKWTWCTMCWKCTTYWMAGTMMYRYRR 101

Qy        328 AACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACC 387
               :::  :|:: ::||     :|: :: ::| :::: : :   | ::: ::|  ::  ::
Db        100 YWYYAKRAKWSKRCTWSTTCYCMKYMAKKCWSYWWSMSMMKWGKSMWWKWTYYYYYYMMK 41

Qy        388 GCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTT 426
              :   :    :|   : ::|  : : ::   |:|: |::
Db         40 WSKMTYWSMMSCYARKCWRTYAKTYTWMTCMTWGKRTYK 2
```

RESULT 7
US-09-489-039A-2755
; Sequence 2755, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2755
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2755

Query Match 7.0%; Score 32; DB 4; Length 1152;
Best Local Similarity 56.7%; Pred. No. 0.43;
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCC 214
Db 656 CCACGGCCTCGGTGCTGGCTCTGCTCAAGCGGCTGCGTGATGAGTACCAGCTGGCGATCG 715

Qy 215 GTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGAT 258
Db 716 TCCTGATTACCCATGAAATGGACGCGGTGCGCACGGCGGCGGAT 759

RESULT 8
US-09-489-039A-2195/c
; Sequence 2195, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2195
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2195

Query Match 6.9%; Score 31.4; DB 4; Length 534;
Best Local Similarity 52.7%; Pred. No. 0.44;
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGT 133
Db 501 AAAGAGGCGGCGACGCAGCCGGTAGCGGTGCCTGCTGCACCGTCGGAAGCGACTCAGGGC 442

Qy 134 TGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTA 193
Db 441 GGCTTCTTTAGCCGCTGGCTTCAGGCGCTGTGGACGCTGTTCTGGAACACGATCCCGGTT 382

Qy 194 AATATGATC 202
Db 381 TATATCCTC 373

RESULT 9

US-09-489-039A-2162
; Sequence 2162, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2162
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (702)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-489-039A-2162

Query Match 6.9%; Score 31.4; DB 4; Length 1818;
Best Local Similarity 52.7%; Pred. No. 0.92;
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGT 133
Db 1363 AAAGAGGCGGCGACGCAGCCGGTAGCGGTGCCTGCTGCACCGTCGGAAGCGACTCAGGGC 1422

Qy 134 TGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTA 193
Db 1423 GGCTTCTTTAGCCGCTGGCTTCAGGCGCTGTGGACGCTGTTCTGGAACACGATCCCGGTT 1482

Qy 194 AATATGATC 202
Db 1483 TATATCCTC 1491

RESULT 10
US-09-252-991A-14225/c
; Sequence 14225, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14225
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14225

Query Match 6.8%; Score 31.2; DB 4; Length 549;
Best Local Similarity 53.2%; Pred. No. 0.53;
Matches 66; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
Db 282 CGTTGCACCGTCAGGCTCGGGGTCAGGCCGGTGCTGCACTGGATCTGGATATTGCCGCTG 223

Qy 191 GTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGG 250
Db 222 GTGCCGACGGTCTGCGCGTCGACCACGTTGGTAAGGTCGGAATGGCTGCCGAAGTCCAGG 163

```
Qy        251 GTGC 254
              |||| |
Db        162 GTTC 159


RESULT 11
US-09-252-991A-14137/c
; Sequence 14137, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 14137
;   LENGTH: 696
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14137

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 696;
  Best Local Similarity   53.2%;  Pred. No. 0.61;
  Matches   66;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
              ||||| |  | |       |||  ||   ||||| || | |||| | |  |  ||
Db        377 CGTTGCACCGTCAGGCTCGGGGTCAGGCCGGTGCTGCACTGGATCTGGATATTGCCGCTG 318

Qy        191 GTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGG 250
              ||    | | ||  ||  ||   ||  |||| ||     | ||||||    | || |||||
Db        317 GTGCCGACGGTCTGCGCGTCGACCACGTTGGTAAGGTCGGAATGGCTGCCGAAGTCCAGG 258

Qy        251 GTGC 254
              |||| |
Db        257 GTTC 254


RESULT 12
US-08-443-104-5/c
; Sequence 5, Application US/08443104
; Patent No. 5691162
;  GENERAL INFORMATION:
;    APPLICANT:  Shuster, Jeffrey R.
;    APPLICANT:  Madden, Mark
;    APPLICANT:  Moyer, Donna L.
;    APPLICANT:  Fuglsang, Claus
;    APPLICANT:  Branner, Sven
;    TITLE OF INVENTION:  NOVEL METALLOPROTEASE HAVING INCREASED
;    TITLE OF INVENTION:  ACTIVITY
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5691162o No. 5691162disk of No. 5691162th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Tape
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/443,104
;      FILING DATE:  17-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/398,489
;      FILING DATE:  03=MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4180.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2052 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-443-104-5

  Query Match             6.8%;  Score 31;  DB 1;  Length 2052;
  Best Local Similarity   45.6%;  Pred. No. 1.4;
  Matches  109;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy        215 GTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              |||   ||| ||  |   | ||  | |    ||  ||| |   | | ||    | | |
Db        648 GTGAGCTTAGCCTCGGGCTTGCTGACGGTACCCTTGGTACCAGTGAAGGTGTAGTGGTTC 589

Qy        275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
              | |||    |   |         || || | || ||  |  ||     || | |  ||
Db        588 TTGCTCTTCTTCTCAGCCTTGGCCTTGTCAGCCTCAACGGGGAGAGAAAGAACATCAACG 529

Qy        335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTG 394
              | |   ||  ||    |||| |    |  || | |||| | |||   |||||    |
Db        528 GTGTCCTTGAGAGCGTCGACGGGGTCTTTCTCGTCACGCTTGGTAAGAGGACCGGGAATC 469

Qy        395 ATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGTAT 453
              | ||  ||||   |   |   || ||    | |  |  || ||  || || | ||||
Db        468 TTGCCCTCGTAGAAGCTGTTTCCGTAGGAGAAGACCTCGCCGTCAGCGCCAATCTGTAT 410


RESULT 13
US-08-238-130-6/c
; Sequence 6, Application US/08238130
; Patent No. 5702934
;  GENERAL INFORMATION:
;    APPLICANT:  Hastrup, Sven
;    APPLICANT:  Branner, Sven
;    APPLICANT:  Jorgensen, Birthe R.
;    APPLICANT:  Christensen, Tove
;    APPLICANT:  Jorgensen, Birgitte B.
;    APPLICANT:  Shuster, Jeffrey R.
;    APPLICANT:  Madden, Mark
;    APPLICANT:  Moyer, Donna L.
;    APPLICANT:  Fuglsang, Claus
;    TITLE OF INVENTION:  PROCESSES FOR PRODUCING AN ENZYME
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5702934o No. 5702934disk of No. 5702934th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Tape
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/238,130
;      FILING DATE:  04-MAY-1994
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 522/93
;      FILING DATE:  05-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  3965.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2052 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..363, 416..802, 856..1821, 1870..2052)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(1..363, 416..802, 856..1821, 1870..2049)
US-08-238-130-6

  Query Match             6.8%;  Score 31;  DB 1;  Length 2052;
  Best Local Similarity   45.6%;  Pred. No. 1.4;
  Matches  109;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy        215 GTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              |||   ||| ||  |   | ||  | |    ||  ||| |   | |  ||    | |  |
Db        648 GTGAGCTTAGCCTCGGGCTTGCTGACGGTACCCTTGGTACCAGTGAAGGTGTAGTGGTTC 589

Qy        275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
              | |||    |     |         || || | || ||  |  ||     || | |  ||
Db        588 TTGCTCTTCTTCTCAGCCTTGGCCTTGTCAGCCTCAACGGGGAGAGAAAGAACATCAACG 529

Qy        335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTG 394
               | |   ||   ||    |||| |     | || | |||| | |||   |||||    |
Db        528 GTGTCCTTGAGAGCGTCGACGGGGTCTTTCTCGTCACGCTTGGTAAGAGGACCGGGAATC 469

Qy        395 ATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGTAT 453
               | ||  ||||   |    |   || ||    | |  |  || ||  || || | ||||
Db        468 TTGCCCTCGTAGAAGCTGTTTCCGTAGGAGAAGACCTCGCCGTCAGCGCCAATCTGTAT 410


RESULT 14
US-08-442-859-5/c
; Sequence 5, Application US/08442859
; Patent No. 5807729
;  GENERAL INFORMATION:
;    APPLICANT:  Shuster, Jeffrey R.
;    APPLICANT:  Madden, Mark
;    APPLICANT:  Moyer, Donna L.
;    APPLICANT:  Fuglsang, Claus
;    APPLICANT:  Branner, Sven
;    TITLE OF INVENTION:  NOVEL METALLOPROTEASE HAVING INCREASED
;    TITLE OF INVENTION:  ACTIVITY
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5807729o No. 5807729disk of No. 5807729th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Tape
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/442,859
;      FILING DATE:  17-MAY-1995
;      CLASSIFICATION:  252
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/398,489
;      FILING DATE:  03=MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4180.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2052 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-442-859-5

  Query Match             6.8%;  Score 31;  DB 1;  Length 2052;
  Best Local Similarity   45.6%;  Pred. No. 1.4;
  Matches  109;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy        215 GTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              |||   ||| ||  |   | ||  | |    ||  ||| |   | |  ||    | |  |
Db        648 GTGAGCTTAGCCTCGGGCTTGCTGACGGTACCCTTGGTACCAGTGAAGGTGTAGTGGTTC 589

Qy        275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
              | |||    |     |         || || | || ||  |  ||     || | |  ||
Db        588 TTGCTCTTCTTCTCAGCCTTGGCCTTGTCAGCCTCAACGGGGAGAGAAAGAACATCAACG 529

Qy        335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTG 394
               | |   ||   ||    |||| |     | || | |||| | |||   |||||    |
Db        528 GTGTCCTTGAGAGCGTCGACGGGGTCTTTCTCGTCACGCTTGGTAAGAGGACCGGGAATC 469

Qy        395 ATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGTAT 453
               | ||  ||||   |    |   || ||    | |  |  || ||  || || | ||||
Db        468 TTGCCCTCGTAGAAGCTGTTTCCGTAGGAGAAGACCTCGCCGTCAGCGCCAATCTGTAT 410


RESULT 15
US-08-398-489-5/c
; Sequence 5, Application US/08398489
; Patent No. 5843753
;  GENERAL INFORMATION:
;    APPLICANT:  Shuster, Jeffrey R.
;    APPLICANT:  Madden, Mark
;    APPLICANT:  Moyer, Donna L.
;    APPLICANT:  Fuglsang, Claus
;    APPLICANT:  Branner, Sven
;    TITLE OF INVENTION:  NOVEL METALLOPROTEASE HAVING INCREASED
;    TITLE OF INVENTION:  ACTIVITY
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5843753o No. 5843753disk of No. 5843753th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Tape
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/398,489
;      FILING DATE:  03-MAR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
```

; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 4180.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-398-489-5

Query Match 6.8%; Score 31; DB 2; Length 2052;
Best Local Similarity 45.6%; Pred. No. 1.4;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

```
Qy        215 GTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAAC 274
              |||   ||| ||  |   | ||  | |    ||  ||| |   | |  ||    | |  |
Db        648 GTGAGCTTAGCCTCGGGCTTGCTGACGGTACCCTTGGTACCAGTGAAGGTGTAGTGGTTC 589

Qy        275 TGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCG 334
              |  |||    |      |         || || | || ||  |  ||     || | |  ||
Db        588 TTGCTCTTCTTCTCAGCCTTGGCCTTGTCAGCCTCAACGGGGAGAGAAAGAACATCAACG 529

Qy        335 ATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTG 394
              | |   ||   ||    |||| |    |  || | |||| | |||   |||||    |
Db        528 GTGTCCTTGAGAGCGTCGACGGGGTCTTTCTCGTCACGCTTGGTAAGAGGACCGGGAATC 469

Qy        395 ATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCACGGCTAACCAGTAT 453
              | ||  ||||   |     |   || ||    | |  |  || ||  || || | ||||
Db        468 TTGCCCTCGTAGAAGCTGTTTCCGTAGGAGAAGACCTCGCCGTCAGCGCCAATCTGTAT 410
```

Search completed: March 18, 2004, 20:20:58
Job time : 62.4971 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	34.6	7.6	1662	14	US-10-156-761-3483	Sequence 3483, Ap
	2	34.6	7.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	3	34	7.5	1344	9	US-09-835-684-6	Sequence 6, Appli
c	4	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appli
c	5	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appli
c	6	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appli
c	7	34	7.5	1344	15	US-10-441-736-5	Sequence 5, Appli
c	8	33.4	7.3	655	14	US-10-184-644-418	Sequence 418, App
c	9	33.4	7.3	655	14	US-10-184-634-418	Sequence 418, App
c	10	33.2	7.3	1848	14	US-10-156-761-7222	Sequence 7222, Ap
c	11	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	12	32.8	7.2	24081	14	US-10-132-134-13	Sequence 13, Appl
	13	32.8	7.2	52101	14	US-10-132-134-1	Sequence 1, Appli
c	14	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
	15	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
	16	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
	17	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
c	18	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
c	19	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appli
c	20	31.4	6.9	460	15	US-10-260-238-787	Sequence 787, App
	21	31.4	6.9	7374	10	US-09-764-891-10207	Sequence 10207, A
	22	31.4	6.9	7374	14	US-10-205-428-1006	Sequence 1006, Ap
c	23	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
	24	31.2	6.8	2266	15	US-10-108-260A-375	Sequence 375, App
	25	31	6.8	1049	12	US-10-142-426-358	Sequence 358, App
	26	31	6.8	1049	14	US-10-123-155-358	Sequence 358, App
	27	31	6.8	1049	14	US-10-146-731-358	Sequence 358, App
	28	31	6.8	1049	14	US-10-140-472-358	Sequence 358, App
	29	31	6.8	1049	14	US-10-141-761-358	Sequence 358, App
	30	31	6.8	1049	14	US-10-142-885-358	Sequence 358, App
	31	31	6.8	1049	14	US-10-158-790-358	Sequence 358, App
	32	31	6.8	1049	15	US-10-137-871-358	Sequence 358, App
	33	31	6.8	1049	15	US-10-140-923-358	Sequence 358, App
	34	31	6.8	1049	15	US-10-141-756-358	Sequence 358, App
	35	31	6.8	1049	15	US-10-141-759-358	Sequence 358, App
	36	31	6.8	1049	15	US-10-140-805-358	Sequence 358, App
	37	31	6.8	1049	15	US-10-140-864-358	Sequence 358, App
c	38	31	6.8	1841	15	US-10-369-493-45479	Sequence 45479, A
c	39	30.8	6.8	524	14	US-10-029-386-4400	Sequence 4400, Ap
	40	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
c	41	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
	42	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
	43	30.6	6.7	1421	12	US-10-425-114-12527	Sequence 12527, A
c	44	30.6	6.7	1573	9	US-09-771-161A-1	Sequence 1, Appli
	45	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1662)
US-10-156-761-3483

Query Match 7.6%; Score 34.6; DB 14; Length 1662;
Best Local Similarity 49.7%; Pred. No. 0.79;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         590 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 531
```

```
Qy        110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
              |  ||| ||  | | |   ||  || | |     | || |||  | | |||
Db        530 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 471

Qy        170 TTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCC 226
              |||  || |   |||  | | |   | ||  |   | |||  |  | | |||
Db        470 TTGAACTTCTTGGCGGCGACGGCGGCGAACGTCTCGTACGTGCCCTGGAGGATGCCC 414
```

RESULT 2
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 7.6%; Score 34.6; DB 14; Length 9025608;
Best Local Similarity 49.7%; Pred. No. 77;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy          50 GTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCA 109
               | || | ||  | ||||| | || |  | | || || || | ||| | ||||||   |
Db     4331781 GCGCGCGGGTCGACGTCGATACAGATCGCGACGCCGTCGTTCATCGTCACGGCGAGCGGC 4331840

Qy         110 ATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGC 169
               |  ||| ||  | | |   ||  || | |     | || |||  | | |||
Db     4331841 TGGGCGCCGCCCATGCCGCCGAGTCCGGCGGTGAGGGTGATCGTGCCGGCGAGGGTGCCG 4331900

Qy         170 TTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCC 226
               |||  || |   |||  | | |   | ||  |   | |||  |  | | |||
Db     4331901 TTGAACTTCTTGGCGGCGACGGCGGCGAACGTCTCGTACGTGCCCTGGAGGATGCCC 4331957
```

RESULT 3
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Dewen
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; TITLE OF INVENTION: DESICCATION
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1.1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy          52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
               |||| || || ||   | ||  | || ||||   || || |  ||||  |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy         112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
               | ||  | |     |||||  | |  ||    |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy         172 GC 173
               ||
Db         89 GC 88
```

RESULT 4
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1.1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy          52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
               |||| || || ||   | ||  | || ||||   || || |  ||||  |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy         112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
               | ||  | |     |||||  | |  ||    |     |   ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy         172 GC 173
               ||
Db         89 GC 88
```

RESULT 5
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF

```
;  FILE REFERENCE: 21829/81
;  CURRENT APPLICATION NUMBER: US/09/879,248
;  CURRENT FILING DATE:  2001-06-12
;  PRIOR APPLICATION NUMBER: 60/212,211
;  PRIOR FILING DATE: 2000-06-16
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-09-879-248-5

  Query Match             7.5%;  Score 34;  DB 9;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 1.1;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 6
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Leon, Ernesto
;  APPLICANT:  Oviedo, Agustin
;  TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
;  TITLE OF INVENTION:  FROM ORNAMENTAL PLANTS
;  FILE REFERENCE: 21829/111
;  CURRENT APPLICATION NUMBER: US/10/010,390
;  CURRENT FILING DATE:  2001-11-05
;  PRIOR APPLICATION NUMBER: 60/248,169
;  PRIOR FILING DATE: 2000-11-13
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-010-390-6

  Query Match             7.5%;  Score 34;  DB 14;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 1.1;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 7
US-10-441-736-5/c
; Sequence 5, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Schading, Richard L.
;  TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;  TITLE OF INVENTION:  RESISTANCE
;  FILE REFERENCE: 21829/203 (EBC-003)
;  CURRENT APPLICATION NUMBER: US/10/441,736
;  CURRENT FILING DATE:  2003-05-20
;  PRIOR APPLICATION NUMBER: 60/107,243
;  PRIOR FILING DATE: 1998-11-05
;  PRIOR APPLICATION NUMBER: 09/431,614
;  PRIOR FILING DATE: 1999-11-02
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Erwinia amylovora
US-10-441-736-5

  Query Match             7.5%;  Score 34;  DB 15;  Length 1344;
  Best Local Similarity   54.9%;  Pred. No. 1.1;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         52 GCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAAT 111
              |||| || || ||   | ||  | || ||||   || || |   |||| |   || ||||
Db        209 GCTCCGGTTGCCGCATTACCTGATTGTGGCGATAGCAGTGACTTTAACAGTTCCGCCAAT 150

Qy        112 AGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
              | ||  | |     ||||| | |  ||   |     |    ||| | | ||||| |   ||
Db        149 AATTGAGCCATTTGCTCAATGGTTTGCCGATCGATGGGTTGTTGCCCCAACGCAGAATTT 90

Qy        172 GC 173
              ||
Db         89 GC 88


RESULT 8
US-10-184-644-418/c
; Sequence 418, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C227
;  CURRENT APPLICATION NUMBER: US/10/184,644
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 418
;   LENGTH: 655
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-644-418

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 655;
  Best Local Similarity   14.9%;  Pred. No. 1.2;
  Matches   39;  Conservative   76;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         18 GGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATG 77
              :::  |||  :   :  ::::    : ::      : ::  |: ::  : : |      :
```

```
Db        325 SSM.CCAT.KS...M..DYDD.C..DNBB..N..AMNBH.MCYTSSANH.BAC...SS.B 266

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137

Db        265 M.YBSTSSNS.ASNKCM.BRASS.B.DA.KD..Y...TAC.S.SASSS.S..ACGBNHA. 206

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATA 197

Db        205 MAC.M.BB.N.S.GHARTMGMKBSS.BHS.ACYTG.MGYBH.A.ACMMYBCWCB.GBAD. 146

Qy        198 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGA 257

Db        145 TGATNS..KC..GAG.YHA.TTG..M..K.W.G..SGHDGM..CN.DYA.ACB.GSSGBY 86

Qy        258 TAATAGTACTATTGAACTGACT 279

Db         85 .SANAGD.SBRACT.GMHGA.B 64


RESULT 9
US-10-184-634-418/c
; Sequence 418, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C217
;  CURRENT APPLICATION NUMBER: US/10/184,634
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 418
;  LENGTH: 655
;  TYPE: PRT
;  ORGANISM: Homo Sapien
US-10-184-634-418

  Query Match             7.3%;  Score 33.4;  DB 14;  Length 655;
  Best Local Similarity   14.9%;  Pred. No. 1.2;
  Matches   39;  Conservative   76;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         18 GGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATG 77

Db        325 SSM.CCAT.KS...M..DYDD.C..DNBB..N..AMNBH.MCYTSSANH.BAC...SS.B 266

Qy         78 GGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAG 137

Db        265 M.YBSTSSNS.ASNKCM.BRASS.B.DA.KD..Y...TAC.S.SASSS.S..ACGBNHA. 206

Qy        138 CATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATA 197

Db        205 MAC.M.BB.N.S.GHARTMGMKBSS.BHS.ACYTG.MGYBH.A.ACMMYBCWCB.GBAD. 146

Qy        198 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGA 257

Db        145 TGATNS..KC..GAG.YHA.TTG..M..K.W.G..SGHDGM..CN.DYA.ACB.GSSGBY 86

Qy        258 TAATAGTACTATTGAACTGACT 279

Db         85 .SANAGD.SBRACT.GMHGA.B 64


RESULT 10
US-10-156-761-7222/c
; Sequence 7222, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7222
;  LENGTH: 1848
;  TYPE: DNA
;  ORGANISM: Streptomyces avermitilis
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(1848)
US-10-156-761-7222

  Query Match             7.3%;  Score 33.2;  DB 14;  Length 1848;
  Best Local Similarity   47.6%;  Pred. No. 2.5;
  Matches   98;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105

Db        834 GGCGGCGTACACCGCGGCAGCGAGGCCGCCGGGTCCGGAGCCGACGATGAGGACGTCGCC 775

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165

Db        774 GTGACGGTTTCCGGCGGGGAGTGACGGCAGGCCGATGAGCCGGGCCAGTTCCGCGTTGCT 715

Qy        166 GCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACC 225

Db        714 GGGGTTGCGCAACAGGGTGGTGTCCCGCCAGATGACGAGCGGTGTCTCCTCCGGGCCGAC 655

Qy        226 CATGAAATGGCACATGCAGGCCAGGG 251

Db        654 GCCGAAGCGGCGCAGCAGGTCCTCGG 629


RESULT 11
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
```

```
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             7.3%;  Score 33.2;  DB 14;  Length 9025608;
  Best Local Similarity   47.6%;  Pred. No. 1.9e+02;
  Matches   98;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105
              ||| | |  |     |||  ||   |     ||  ||| | ||   || |  ||| || |
Db    8638225 GGCGGCGTACACCGCGGCAGCGAGGCCGCCGGGTCCGGAGCCGACGATGAGGACGTCGCC 8638166

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165
              |  |    |||||||  ||| | |  |  |  |  | | | |  |   | |      |||
Db    8638165 GTGACGGTTTCCGGCGGGGAGTGACGGCAGGCCGATGAGCCGGGCCAGTTCCGCGTTGCT 8638106

Qy        166 GCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACC 225
              | | |||| |  ||  | | || ||     | ||||  ||| |  |  ||      |    |
Db    8638105 GGGGTTGCGCAACAGGGTGGTGTCCCGCCAGATGACGAGCGGTGTCTCCTCCGGGCCGAC 8638046

Qy        226 CATGAAATGGCACATGCAGGCCAGGG 251
                |||  ||| ||    | ||  ||
Db    8638045 GCCGAAGCGGCGCAGCAGGTCCTCGG 8638020


RESULT 12
US-10-132-134-13
; Sequence 13, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Yang, Xianshu
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
;  FILE REFERENCE: 3012-2US
;  CURRENT APPLICATION NUMBER: US/10/132,134
;  CURRENT FILING DATE:  2002-04-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
;   LENGTH: 24081
;   TYPE: DNA
;   ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-13

  Query Match             7.2%;  Score 32.8;  DB 14;  Length 24081;
  Best Local Similarity   50.0%;  Pred. No. 14;
  Matches   82;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105
              ||| |  | ||||  | || |   || |  |||| || |||||| |         | || |
Db      20227 GGCCGCACACTGGTGGTCGGCTGGCCCCTGTGGGCCGACGGCGGCATGACGGTGGACGCC 20286

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165
              | ||        | ||| | ||    |   |  ||||     | |  | |  || |  ||| |
Db      20287 GCCACCGAACGCCGCCTGCACGACAGCGTCGGCATGGTGCCGATCCGCGCCCCGCACGGT 20346

Qy        166 GCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGT 209
              | |   || |||| | |||     | | |      || | ||  ||
Db      20347 GTGGAGGCGCTGCTACGCGCCTACGGCACCGGCGACCCGCACGT 20390


RESULT 13
US-10-132-134-1
; Sequence 1, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Yang, Xianshu
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
;  FILE REFERENCE: 3012-2US
;  CURRENT APPLICATION NUMBER: US/10/132,134
;  CURRENT FILING DATE:  2002-04-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 52101
;   TYPE: DNA
;   ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-1

  Query Match             7.2%;  Score 32.8;  DB 14;  Length 52101;
  Best Local Similarity   50.0%;  Pred. No. 21;
  Matches   82;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy         46 GGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGC 105
              ||| |  | ||||  | || |   || |  |||| || |||||| |         | || |
Db      37860 GGCCGCACACTGGTGGTCGGCTGGCCCCTGTGGGCCGACGGCGGCATGACGGTGGACGCC 37919

Qy        106 GGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCT 165
              | ||        | ||| | ||    |   |  ||||     | |  | |  || |  ||| |
Db      37920 GCCACCGAACGCCGCCTGCACGACAGCGTCGGCATGGTGCCGATCCGCGCCCCGCACGGT 37979

Qy        166 GCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGT 209
              | |   || |||| | |||     | | |      || | ||  ||
Db      37980 GTGGAGGCGCTGCTACGCGCCTACGGCACCGGCGACCCGCACGT 38023


RESULT 14
US-10-369-493-33133/c
; Sequence 33133, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33133
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Xylella fastidiosa
US-10-369-493-33133

  Query Match             7.1%;  Score 32.2;  DB 15;  Length 1083;
  Best Local Similarity   59.1%;  Pred. No. 4.1;
  Matches   55;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTG 421
              ||  | ||| |||||  |  | |||||     ||  || | ||| | ||| ||||
Db        488 ATCTCACCGAGCTGGGATAAAAAACCGCGCACAATCACACCATAACGCTGCTTCAGCCAC 429

Qy        422 GTTTTGGCAACAACGCCACGGCTAACCAGTATT 454
                |||| ||||| || |||    | ||  | |||
Db        428 TTTTTAGCAACCACCCCAGCAGTCACACGCATT 396


RESULT 15
US-09-815-242-6028
; Sequence 6028, Application US/09815242
```

```
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6028
;   LENGTH: 930
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(930)
US-09-815-242-6028

  Query Match             7.0%;  Score 31.8;  DB 9;  Length 930;
  Best Local Similarity   52.7%;  Pred. No. 5.2;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              | |||  |  || ||||    | |  || | | ||| ||   |   ||  |  || |||
Db        233 AGCAGCAGGGCGGTAAAGCACCGGCAATGATGGCCGGTCACAGCCTGGGGGAATACTCCG 292

Qy        371 CGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCA 430
              ||||||||          |    ||||| |   |     |||||||| |||||    |  ||
Db        293 CGCTGGTTTGCGCTGGTGTGATTGATTTCGCTGATGCGGTGCGTCTGGTTGAGATGCGCG 352

Qy        431 ACAACGCCACG 441
               |||   || |
Db        353 GCAAGTTCATG 363


Search completed: March 19, 2004, 03:50:20
Job time : 276.734 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-19	Sequence 19, Appl
2	412.8	90.5	456	6	US-08-233-642A-56	Sequence 56, Appl
3	412.8	90.5	456	23	US-09-543-407-1	Sequence 1, Appli
4	385.6	84.6	456	23	US-09-543-407-25	Sequence 25, Appl
5	363.2	79.6	456	23	US-09-543-407-17	Sequence 17, Appl
6	361.6	79.3	456	23	US-09-543-407-11	Sequence 11, Appl
7	361.6	79.3	456	23	US-09-543-407-13	Sequence 13, Appl
8	361.6	79.3	456	23	US-09-543-407-23	Sequence 23, Appl
9	360	78.9	456	23	US-09-543-407-15	Sequence 15, Appl
10	360	78.9	456	23	US-09-543-407-27	Sequence 27, Appl
11	358.4	78.6	456	23	US-09-543-407-21	Sequence 21, Appl
12	355.2	77.9	456	23	US-09-543-407-29	Sequence 29, Appl
13	288.6	63.3	361	6	US-08-233-642A-54	Sequence 54, Appl
14	236.8	51.9	456	23	US-09-543-407-3	Sequence 3, Appli
15	236.8	51.9	675	13	US-08-978-878-3	Sequence 3, Appli
16	236.8	51.9	675	31	US-09-741-873B-3	Sequence 3, Appli
17	234.4	51.4	453	92	US-60-352-946-1	Sequence 1, Appli
18	234.4	51.4	453	101	US-60-444-371-1	Sequence 1, Appli
19	227.2	49.8	477	17	US-09-252-691-172	Sequence 172, App
20	227.2	49.8	477	17	US-09-252-691C-172	Sequence 172, App
21	227.2	49.8	477	51	US-10-417-886-172	Sequence 172, App
22	192.8	42.3	396	13	US-08-978-878-1	Sequence 1, Appli
23	168	36.8	360	31	US-09-741-873B-1	Sequence 1, Appli
24	64.8	14.2	522	17	US-09-252-691-170	Sequence 170, App
25	64.8	14.2	522	17	US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	51	US-10-417-886-170	Sequence 170, App
27	51.2	11.2	100	45	US-10-146-492B-77	Sequence 77, Appl
C 28	50	11.0	78	23	US-09-543-407-52	Sequence 52, Appl
29	50	11.0	78	23	US-09-543-407-53	Sequence 53, Appl
30	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appli
31	45	9.9	78	23	US-09-543-407-49	Sequence 49, Appl
C 32	43.4	9.5	78	23	US-09-543-407-48	Sequence 48, Appl
33	42.4	9.3	100	45	US-10-146-492B-78	Sequence 78, Appl
C 34	42	9.2	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
C 35	36.8	8.1	1290	37	US-09-902-540-5006	Sequence 5006, Ap
C 36	36.8	8.1	12950	37	US-09-902-540-1036	Sequence 1036, Ap
C 37	36.2	7.9	632	33	US-09-815-264-23539	Sequence 23539, A
C 38	36.2	7.9	805	26	US-09-620-392-61621	Sequence 61621, A
C 39	36.2	7.9	805	30	US-09-702-134-34173	Sequence 34173, A
C 40	36.2	7.9	805	33	US-09-815-264-61323	Sequence 61323, A
41	35.8	7.9	387	11	US-08-790-774-9199	Sequence 9199, Ap
42	35.8	7.9	387	37	US-09-912-293-217377	Sequence 217377,
C 43	35.8	7.9	1281	20	US-09-417-507-7264	Sequence 7264, Ap
C 44	35	7.7	2256646	52	US-10-470-565-1	Sequence 1, Appli
C 45	34.8	7.6	542	24	US-09-565-306-14113	Sequence 14113, A

ALIGNMENTS

```
RESULT 1
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-136;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
```

```
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-642A-56

  Query Match             90.5%;  Score 412.8;  DB 6;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 2.1e-122;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||||| |||    |    ||||||      | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 3
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match             90.5%;  Score 412.8;  DB 23;  Length 456;
  Best Local Similarity   94.1%;  Pred. No. 2.1e-122;
  Matches  429;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              ||||||||||||||||| |||    |    ||||||      | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 4
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             84.6%;  Score 385.6;  DB 23;  Length 456;
  Best Local Similarity   90.4%;  Pred. No. 1.4e-113;
  Matches  412;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
```

```
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |  ||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTTATGATCAG 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
                | | |||||     |   ||    | | ||  | |    || |  ||
Db        181 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 5
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 79.6%; Score 363.2; DB 23; Length 456;
Best Local Similarity 87.3%; Pred. No. 2.5e-106;
Matches 398; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||        |   ||   |   | | ||  | |    || | |||  ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 8.2e-106;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |   ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              |||||||||||||||||||||| |           | | || | ||  |   | | | |
Db        361 AATAACGCCGCGCTGGTTAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 420
```

```
Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |    ||  ||||||||||||||||||||||||||||
Db        421 GCACATGCAAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             79.3%;  Score 361.6;  DB 23;  Length 456;
  Best Local Similarity   87.1%;  Pred. No. 8.2e-106;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                  ||    | ||||  ||        |  || | ||||||||||||||||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             79.3%;  Score 361.6;  DB 23;  Length 456;
  Best Local Similarity   87.1%;  Pred. No. 8.2e-106;
  Matches  397;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |    ||||||     | |     | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              |||||||||||||||||||||||||||||||||  |||     ||  ||     |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                || |    || |  | ||||||||||||||||||||||||||||||||||||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match           78.9%;  Score 360;  DB 23;  Length 456;
  Best Local Similarity  86.8%;  Pred. No. 2.7e-105;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||  |   ||  ||  | | | |   | |   |  ||  ||||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |  ||||||  | |  | |  ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 10
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match           78.9%;  Score 360;  DB 23;  Length 456;
  Best Local Similarity  86.8%;  Pred. No. 2.7e-105;
  Matches  396;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |||||||||||||||| |||   |  ||||||  | |  | |  ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match           78.6%;  Score 358.4;  DB 23;  Length 456;
  Best Local Similarity  86.6%;  Pred. No. 8.9e-105;
  Matches  395;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
                |||||||||||||||| |||   |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | ||||||||||||||||||  | |      | ||  |    ||  |  |    | |
Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
                ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
                |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

  Query Match             77.9%;  Score 355.2;  DB 23;  Length 456;
  Best Local Similarity   86.2%;  Pred. No. 9.7e-104;
  Matches  393;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
                ||||||||||||||| |||   |    ||||||     | |      | |    ||  ||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
                | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
                ||||||   || |||  |   |                |     | |  ||    ||||||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
                |||||||||||||||||||||||||||||||||||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-642A-54

  Query Match             63.3%;  Score 288.6;  DB 6;  Length 361;
  Best Local Similarity   91.3%;  Pred. No. 3.4e-82;
  Matches  306;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         64 GTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
                |||||  |||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 183
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120
```

```
Qy        184 GATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243
              ||||||||||||||| |||   |   ||||||     | |     | |    ||  ||| |
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 398
              |||||||||||||||||||||||||||||||||||
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-3

  Query Match             51.9%;  Score 236.8;  DB 23;  Length 456;
  Best Local Similarity   70.0%;  Pred. No. 2.3e-65;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  ||||||||||| | |||   |    ||||||    ||     | | |    ||| ||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| ||||||  ||| |  |||  |||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

```
RESULT 15
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
;  APPLICANT: NORMARK, Staffan
;  APPLICANT: OLSEN, Arne
;  TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
;  FILE REFERENCE: 012889-081
;  CURRENT APPLICATION NUMBER: US/08/978,878
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: SE 8801723-1
;  EARLIER FILING DATE: 1988-05-06
;  EARLIER APPLICATION NUMBER: US 07/347,189
;  EARLIER FILING DATE: 1989-05-04
;  EARLIER APPLICATION NUMBER: US 07/789,437
;  EARLIER FILING DATE: 1991-11-06
;  EARLIER APPLICATION NUMBER: US 07/970,846
;  EARLIER FILING DATE: 1992-11-03
;  EARLIER APPLICATION NUMBER: US 08/187,865
;  EARLIER FILING DATE: 1994-01-28
;  EARLIER APPLICATION NUMBER: US 08/318,519
;  EARLIER FILING DATE: 1994-10-05
;  EARLIER APPLICATION NUMBER: US 08/495,959
;  EARLIER FILING DATE: 1995-06-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-08-978-878-3

  Query Match             51.9%;  Score 236.8;  DB 13;  Length 675;
  Best Local Similarity   70.0%;  Pred. No. 2.6e-65;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  ||||||||||| | |||   |    ||||||    ||     | | |    ||| ||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || |||||  ||   ||| ||| | || |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| ||||||  ||| |  |||  |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

Search completed: March 19, 2004, 03:01:58
Job time : 2652.77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	236.8	51.9	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	192.8	42.3	396	5	US-09-741-873C-1	Sequence 1, Appli
c	3	30.8	6.8	1575	6	US-10-788-792-27	Sequence 27, Appl
	4	30.4	6.7	848	6	US-10-100-683-643	Sequence 643, App
	5	29.4	6.4	300	6	US-10-767-701-28437	Sequence 28437, A
c	6	29.2	6.4	249	6	US-10-771-241-117	Sequence 117, App
	7	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
c	8	29	6.4	1046	6	US-10-048-770C-3	Sequence 3, Appli
c	9	28.8	6.3	3816	1	PCT-US04-04280-78	Sequence 78, Appl
c	10	28.8	6.3	4333	6	US-10-152-372-233	Sequence 233, App
c	11	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
	12	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
c	13	28.6	6.3	1295	6	US-10-100-683-5480	Sequence 5480, Ap
c	14	28.6	6.3	3416	6	US-10-100-683-3397	Sequence 3397, Ap
c	15	28.6	6.3	8206	6	US-10-021-698A-3529	Sequence 3529, Ap
	16	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	17	28.2	6.2	201	7	US-60-548-091-22688	Sequence 22688, A
	18	28.2	6.2	596	6	US-10-767-701-5378	Sequence 5378, Ap
c	19	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
c	20	28.2	6.2	1746	1	PCT-US04-05654-2555	Sequence 2555, Ap
c	21	28.2	6.2	14449	6	US-10-021-698A-924	Sequence 924, App
	22	28.2	6.2	225724	6	US-10-021-698A-711	Sequence 711, App
	23	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
c	24	28	6.1	554	6	US-10-767-701-27293	Sequence 27293, A
	25	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
c	26	27.8	6.1	1400	7	US-60-545-213-2134	Sequence 2134, Ap
c	27	27.8	6.1	1400	7	US-60-545-213-2135	Sequence 2135, Ap
c	28	27.8	6.1	1400	7	US-60-545-213-6406	Sequence 6406, Ap
c	29	27.8	6.1	1400	7	US-60-545-213-6407	Sequence 6407, Ap
	30	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
c	31	27.6	6.1	363	6	US-10-767-701-29572	Sequence 29572, A
	32	27.6	6.1	463	6	US-10-767-701-17923	Sequence 17923, A
	33	27.6	6.1	529	6	US-10-767-701-19620	Sequence 19620, A
	34	27.6	6.1	551	6	US-10-767-701-3483	Sequence 3483, Ap
	35	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
	36	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
	37	27.6	6.1	617	6	US-10-781-469-57	Sequence 57, Appl
	38	27.6	6.1	632	6	US-10-100-683-3854	Sequence 3854, Ap
c	39	27.6	6.1	946	6	US-10-784-004-232	Sequence 232, App
	40	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
c	41	27.6	6.1	1818	6	US-10-773-236-118	Sequence 118, App
	42	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
	43	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
	44	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	45	27.6	6.1	8654	6	US-10-021-698A-3499	Sequence 3499, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 51.9%; Score 236.8; DB 5; Length 675;
Best Local Similarity 70.0%; Pred. No. 3.4e-72;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              ||||||||||||||||| |||||| | |||||||||||| | || || || ||| ||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  |||||||||||| | |||    |     ||||||     ||      | |      ||| ||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || |||||  ||    ||| ||| | ||| |
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| ||||||  | |||  ||| |  ||| |||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538


RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873C
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873C-1

  Query Match             42.3%;  Score 192.8;  DB 5;  Length 396;
  Best Local Similarity   67.9%;  Pred. No. 4.5e-57;
  Matches  269;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |   ||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  |||||||||||| | |||    |     ||||||     ||      | |      ||| ||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || |||||  ||    ||| ||| | || |

Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||     || | ||| |||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| ||||||  | |||  ||| |  ||| |||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396


RESULT 3
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
;  APPLICANT: Bayer Pharmaceuticals Corporation
;  APPLICANT:  Eveleigh, Deepa
;  APPLICANT:  Bigwood, Douglas
;  TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
;  FILE REFERENCE: 5152
;  CURRENT APPLICATION NUMBER: US/10/788,792
;  CURRENT FILING DATE:  2004-02-27
;  PRIOR APPLICATION NUMBER: US 60/450,655
;  PRIOR FILING DATE: 2003-02-28
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
;   LENGTH: 1575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-788-792-27

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 1575;
  Best Local Similarity   55.7%;  Pred. No. 1.2;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        317 GGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGG 376
              |||  | | | | || ||  | | || |  || | |||    || |||| ||||||||
Db        688 GGATTTCCATACAGTCTGACCGTGCAGTTGCTCAGTTCGGTCAAAAGAACGTCGCGCTGG 629

Qy        377 TTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGG 422
              | | |   | |  |||  |  |    |   |||  |  ||||||||
Db        628 TGCAACTCGCTGGCTCTCTTCTCCAAGACTTGGGACTCCAGGTTGG 583


RESULT 4
US-10-100-683-643
; Sequence 643, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
```

```
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 643
;   LENGTH: 848
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-643

  Query Match             6.7%;  Score 30.4;  DB 6;  Length 848;
  Best Local Similarity   49.4%;  Pred. No. 1.2;
  Matches   79;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              ||| ||  | | ||  | ||   |  |||||     |       |   ||  |||
Db          1 GGCACGAGCACTACTGTAAGAGCTGGTCAGTGAATGTGGTTGCAGCATGGCCTTTGGGCA 60

Qy        178 CAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 237
              |||  |  |   ||| ||  | |||||||||  ||    |   |   || ||| ||
Db         61 AGAAGTAACCCATTTAACTAAAACCAGCTGGTTGGCCCCACTCAGATTTATCAAAGGGTT 120

Qy        238 CATGCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGA 277
                ||    || ||| | ||||| |  |  ||||  ||| |
Db        121 ACTGGGTCCCTGGGGGTGGATATTGCTTATATTAGACTTA 160


RESULT 5
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

  Query Match             6.4%;  Score 29.4;  DB 6;  Length 300;
  Best Local Similarity   54.1%;  Pred. No. 1.6;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         14 AAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCAC 73
              |||  |  |  | || |||     |        ||||  | |      | |  || || ||| |
Db        123 AAGAAGGCGGTTACGAAGCCGGTGGTTACGGTGGCGGCGGCGGCGGTTACGGCGGTCCCC 182

Qy         74 AATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGG 124
                |||||||||||||| ||     |||||||||  |  ||  |  || ||
Db        183 GTAAGGGCGGCGGCGGTTATGGCGGCGGCGGCGGTTACGGTGGCAACCGGG 233


RESULT 6
US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-117

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 249;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||       || || ||| |    | |  | |   || | | |||
Db        180 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 121

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    |  |  ||  || || || ||| |  | |||    |||| | | |    || |
Db        120 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 61

Qy        415 CA 416
              |
Db         60 GA 59


RESULT 7
US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 757
;   TYPE: DNA
;   ORGANISM: E. Coli
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(757)
;   OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 757;
  Best Local Similarity   52.5%;  Pred. No. 2.9;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        295 AATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATAC 354
              || | ||   ||||||||       || || ||| |    | |  | |   || | | |||
Db        264 AAAATTGAAGCCATCGACTTTACCAAAGCCAAACAGCGTGCTGATCCATCCGCCGACTAC 323

Qy        355 GGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGT 414
              |||    |  |  ||  || || || ||| |  | |||    |||| | | |    || |
Db        324 GGCACCGACTATCCCCAGCAGGATAGTCATAAAGAATCCACCTCCATCTTTACCTGGCAT 383

Qy        415 CA 416
              |
```

```
Db        384 GA 385


RESULT 8
US-10-048-770C-3/c
; Sequence 3, Application US/10048770C
; GENERAL INFORMATION:
;  APPLICANT: NEW BIOTECHNIC, S.A.
;  APPLICANT:  UNIVERSIDAD DE SEVILLA
;  TITLE OF INVENTION: ANTI-FUNGAL COMPOSITIONS AND METHOD FOR CONTROLLING FUNGI
;  FILE REFERENCE: 1837-2
;  CURRENT APPLICATION NUMBER: US/10/048,770C
;  CURRENT FILING DATE:  2002-01-31
;  PRIOR APPLICATION NUMBER: PCT/ES00/00292
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 1046
;   TYPE: DNA
;   ORGANISM: Trichoderma harzianum
US-10-048-770C-3

  Query Match             6.4%;  Score 29;  DB 6;  Length 1046;
  Best Local Similarity   49.0%;  Pred. No. 4;
  Matches   77;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGT 216
              || |||| ||    |  |  ||| | | | ||  |||| | | || |   ||||| ||
Db        760 GCAAACGTTGGTGATATCCATCAATGAGGTTCCGCTAAAGACGGTCTGACCGTTACGAGT 701

Qy        217 GTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGAACTG 276
              |  | |    ||||  |||| |||| || |  |||  | | |     |  |
Db        700 CAGAGCAAACTTTTGCTGGCCGATGCTGGCCGGGATTTGGAAGAGATTGGCTCCGGCGTT 641

Qy        277 ACTCAGAATGGTTTCAGAAATAATGCCACCATCGACC 313
                | | |  |||| ||   |    |||     ||||
Db        640 GGCCTGGAACGTTTGAGTGGTGCCACCAGACGTGACC 604


RESULT 9
PCT-US04-04280-78/c
; Sequence 78, Application PC/TUS0404280
; GENERAL INFORMATION:
;  APPLICANT: INCYTE CORPORATION; KABLE, Amy E.;
;  APPLICANT:  YUE, Henry; BAUGHN, Mariah R.;
;  APPLICANT:  TRIBOULEY, Catherine M.; RING, Huijun Z.;
;  APPLICANT:  TRAN, Uyen K.; EMERLING, Brooke E.;
;  APPLICANT:  RAMKUMAR, Jayalaxmi; HAFALIA, April J.A.;
;  APPLICANT:  SWARNAKAR, Anita; LEE, Soo Yeun;
;  APPLICANT:  CHAWLA, Narinder K.; GIETZEN, Kimberly J.;
;  APPLICANT:  MARQUIS, Joseph P.; ELLIOTT, Vicki S.;
;  APPLICANT:  BECHA, Shanya D.; FAVERO, Kristin D.;
;  APPLICANT:  WANG, Jonathan T.; NAIDU, Sangeeta;
;  APPLICANT:  HAWKINS, Phillip R.; JIN, Pei;
;  APPLICANT:  CHIEN, David
;  TITLE OF INVENTION: ENZYMES
;  FILE REFERENCE: PF-1507 PCT
;  CURRENT APPLICATION NUMBER: PCT/US04/04280
;  CURRENT FILING DATE:  2004-02-12
;  PRIOR APPLICATION NUMBER: US 60/447,246
;  PRIOR FILING DATE: 2003-02-12
;  PRIOR APPLICATION NUMBER: US 60/449,087
;  PRIOR FILING DATE: 2003-02-21
;  PRIOR APPLICATION NUMBER: US 60/450,622
;  PRIOR FILING DATE: 2003-02-26
;  PRIOR APPLICATION NUMBER: US 60/456,704
;  PRIOR FILING DATE: 2003-03-21
;  PRIOR APPLICATION NUMBER: US 60/463,194
;  PRIOR FILING DATE: 2003-04-15
;  PRIOR APPLICATION NUMBER: US 60/469,358
;  PRIOR FILING DATE: 2003-05-09
;  PRIOR APPLICATION NUMBER: US 60/475,532
;  PRIOR FILING DATE: 2003-06-02
;  PRIOR APPLICATION NUMBER: US 60/476,278
;  PRIOR FILING DATE: 2003-06-04
;  PRIOR APPLICATION NUMBER: US 60/483,395
;  PRIOR FILING DATE: 2003-06-27
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: PERL Program
; SEQ ID NO 78
;   LENGTH: 3816
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 7517280CB1
PCT-US04-04280-78

  Query Match             6.3%;  Score 28.8;  DB 1;  Length 3816;
  Best Local Similarity   50.7%;  Pred. No. 8.8;
  Matches   69;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              ||||  |||  |    |||| |   | ||  | |  ||  |     ||| ||    |||
Db       2245 GGCCATGACAAAGGTATGAGGAAAGAGCATAGCTGCACCCTTCCTCATGCCCTGAAGCTG 2186

Qy        178 CAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 237
              || || | ||||   || ||||      ||   | | ||| |   | |||| |    |
Db       2185 CACAGTGTTGCCCTGGAACATGACTCCATGCACATCAGTCTCTGTGCCCAGGCCGAGCAG 2126

Qy        238 CATGCAGGCCAGGGTG 253
                  |||||||  |||
Db       2125 GTGCCAGGCCACTGTG 2110


RESULT 10
US-10-152-372-233/c
; Sequence 233, Application US/10152372
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C394
;  CURRENT APPLICATION NUMBER: US/10/152,372
;  CURRENT FILING DATE:  2002-05-21
;  PRIOR APPLICATION NUMBER: 60/049911
;  PRIOR FILING DATE: 1997-06-18
;  PRIOR APPLICATION NUMBER: 60/056974
;  PRIOR FILING DATE: 1997-08-26
;  PRIOR APPLICATION NUMBER: 60/059113
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059115
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059117
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059122
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059184
;  PRIOR FILING DATE: 1997-09-17
```

```
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059352
;  PRIOR FILING DATE: 1997-09-19
;  PRIOR APPLICATION NUMBER: 60/059588
;  PRIOR FILING DATE: 1997-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 233
;   LENGTH: 4333
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-152-372-233

  Query Match             6.3%;  Score 28.8;  DB 6;  Length 4333;
  Best Local Similarity   50.7%;  Pred. No. 9.4;
  Matches   69;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        118 GGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTG 177
              |||| ||| |    |||| |   | ||   | | ||  |     ||| ||    |||
Db       2191 GGCCATGACAAAGGTATGAGGAAAGAGCATAGCTGCACCCTTCCTCATGCCCTGAAGCTG 2132

Qy        178 CAAAGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 237
              || || | |||||   || ||||     ||   | | || |   | |||| |    |
Db       2131 CACAGTGTTGCCCTGGAACATGACTCCATGCACATCAGTCTCTGTGCCCAGGCCGAGCAG 2072

Qy        238 CATGCAGGCCAGGGTG 253
                |||||||  |||
Db       2071 GTGCCAGGCCACTGTG 2056


RESULT 11
US-10-417-375A-143/c
; Sequence 143, Application US/10417375A
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375A
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143
;   LENGTH: 83009
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(83009)
;   OTHER INFORMATION: n = A,T,C or G
US-10-417-375A-143

  Query Match             6.3%;  Score 28.8;  DB 6;  Length 83009;
  Best Local Similarity   48.8%;  Pred. No. 40;
  Matches   78;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        281 AGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTA 340
              |||||  |   ||||   ||   | | || |  ||   |  |   |  | |||  ||||
Db      14056 AGAATATTAAAAGAATGCATAATATCCTCAAAAAGAACATTGACTAGTATTCCTTTATTC 13997

Qy        341 CTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCA 400
              || ||    |       | ||||    |   |     || ||| |  |   |  ||| |
Db      13996 CTTTCAAGTACACTTAAGTTAATGCTTCAATGAAATACAAGCAGGCACCTATTACTTCTA 13937

Qy        401 GCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCAC 440
              |  |||||   | ||||  | | ||   |  || | |||||
Db      13936 GAATAATGAGACCTCAGATGGATTACAGGCACGATGCCAC 13897


RESULT 12
PCT-US04-05654-1030
; Sequence 1030, Application PC/TUS0405654
; GENERAL INFORMATION:
;  APPLICANT: Sherman, Bradley K
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline  E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Dubell III, Arnold N
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Repetti, Peter
;  APPLICANT:  Century, Karen
;  APPLICANT:  Gutterson, Neal
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Kumimoto, Roderick W
;  APPLICANT:  Pilgrim, Marsha L
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 PCT
;  CURRENT APPLICATION NUMBER: PCT/US04/05654
;  CURRENT FILING DATE:  2004-03-03
;  PRIOR APPLICATION NUMBER: 10/374,780
;  PRIOR FILING DATE: 2003-02-25
;  PRIOR APPLICATION NUMBER: 10/675,852
;  PRIOR FILING DATE: 2003-09-30
;  NUMBER OF SEQ ID NOS: 2950
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1030
;   LENGTH: 1186
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
PCT-US04-05654-1030

  Query Match             6.3%;  Score 28.6;  DB 1;  Length 1186;
  Best Local Similarity   59.0%;  Pred. No. 5.9;
  Matches   49;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         44 CTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCG 103
              | || | |||||  || | || ||       | | || ||||||   |  |   ||||
Db        606 CAGGGACTGCTCCAGCGGCCGGCGGCGGCTACGGCGGTGGCGGCGGCGGTGGTGGTGGCG 665

Qy        104 GCGGCAATAGTTCCGGCCCGGAC 126
              || || | |  | ||||| || |
Db        666 GCTGCTACAACTGCGGCCAGGCC 688


RESULT 13
US-10-100-683-5480/c
; Sequence 5480, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
```

```
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5480
;   LENGTH: 1295
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-5480

  Query Match             6.3%;  Score 28.6;  DB 6;  Length 1295;
  Best Local Similarity   49.7%;  Pred. No. 6.1;
  Matches   73;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        242 CAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
              | | ||  | |||   |||     |  | | | ||| ||      ||        |  | ||
Db        767 CGGACCCTGCTGCCCCTAACCTCAAGAGTCAGCTGGCTGCAGCTGCCAGGGCAGTAACTG 708

Qy        302 CCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
               || |||| | |||   | | |||     | || |     |   ||||||  | |   ||
Db        707 ACAGCATCAATCAGCTCATCACTATGTGCACCCAGCAGGCACCCGGCCAGAAGGAGTGTG 648

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCG 388
              ||||||||  || ||      |||| |
Db        647 ATAACGCCCTGCGGGAATTGGAGACGG 621


RESULT 14
US-10-100-683-3397/c
; Sequence 3397, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3397
;   LENGTH: 3416
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-3397

  Query Match             6.3%;  Score 28.6;  DB 6;  Length 3416;
  Best Local Similarity   49.7%;  Pred. No. 9.8;
  Matches   73;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        242 CAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
              | | ||  | |||   |||     |  | | | ||| ||      ||        |  | ||
Db        761 CGGACCCTGCTGCCCCTAACCTCAAGAGTCAGCTGGCTGCAGCTGCCAGGGCAGTAACTG 702

Qy        302 CCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
               || |||| | |||   | | |||     | || |     |   ||||||  | |   ||
Db        701 ACAGCATCAATCAGCTCATCACTATGTGCACCCAGCAGGCACCCGGCCAGAAGGAGTGTG 642

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCG 388
              ||||||||  || ||      |||| |
Db        641 ATAACGCCCTGCGGGAATTGGAGACGG 615


RESULT 15
US-10-021-698A-3529/c
; Sequence 3529, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 3529
;   LENGTH: 8206
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (401)
;   OTHER INFORMATION: a, t, c or g
US-10-021-698A-3529

  Query Match             6.3%;  Score 28.6;  DB 6;  Length 8206;
  Best Local Similarity   49.7%;  Pred. No. 15;
  Matches   73;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        242 CAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
              | | ||  | |||   |||     |  | | | ||| ||      ||        |  | ||
Db       4174 CGGACCCTGCTGCCCCTAACCTCAAGAGTCAGCTGGCTGCAGCTGCCAGGGCAGTAACTG 4115

Qy        302 CCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTA 361
               || |||| | |||   | | |||     | || |     |   ||||||  | |   ||
Db       4114 ACAGCATCAATCAGCTCATCACTATGTGCACCCAGCAGGCACCCGGCCAGAAGGAGTGTG 4055

Qy        362 ATAACGCCGCGCTGGTTAATCAGACCG 388
              ||||||||  || ||      |||| |
Db       4054 ATAACGCCCTGCGGGAATTGGAGACGG 4028


Search completed: March 19, 2004, 03:10:16
Job time : 55.1445 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-19
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtggc.....ccacggctaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	237.4	52.1	680	12	BJ618688	BJ618688 BJ618688
c	2	37.4	8.2	484	9	AV430994	AV430994 AV430994
c	3	35.4	7.8	664	13	CA141136	CA141136 SCJFRT205
	4	34.8	7.6	675	13	BX251172	BX251172 BX251172
c	5	34.6	7.6	699	12	BI959328	BI959328 HVSMEn001
	6	34.2	7.5	296	9	AV058630	AV058630 AV058630
c	7	34.2	7.5	669	28	AZ959284	AZ959284 2M0226G20
	8	34.2	7.5	978	28	BZ569281	BZ569281 pacs2-164
	9	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
c	10	34.2	7.5	1137	28	BZ565783	BZ565783 pacs2-164
c	11	34	7.5	234	10	BB581987	BB581987 BB581987
c	12	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
	13	34	7.5	500	28	AZ180696	AZ180696 SP_0181_A
c	14	33.8	7.4	365	9	AA746477	AA746477 nw62e02.s
c	15	33.8	7.4	558	14	CD374421	CD374421 TNWbmfC2H
c	16	33.8	7.4	559	13	BU765156	BU765156 sas23f11.
c	17	33.8	7.4	902	28	BZ569250	BZ569250 pacs2-164
	18	33.8	7.4	941	29	CG178280	CG178280 PUJDF79TD
c	19	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetraodon
	20	33.6	7.4	680	10	BF668888	BF668888 602123291
	21	33.6	7.4	966	13	BQ718166	BQ718166 AGENCOURT
c	22	33.4	7.3	505	6	AL809250	Al809250 Triticum
	23	33.4	7.3	523	12	BJ334624	BJ334624 BJ334624
	24	33.4	7.3	715	13	BU444683	BU444683 603765715
	25	33.4	7.3	1064	13	BQ939900	BQ939900 AGENCOURT
c	26	33.2	7.3	317	12	BM333324	BM333324 MEST189-B
	27	33.2	7.3	625	10	BF501499	BF501499 AT16857.5
	28	33.2	7.3	629	9	AU295325	AU295325 AU295325
c	29	33.2	7.3	806	13	BX888120	BX888120 BX888120
	30	33	7.2	509	12	BM896625	BM896625 ph53g06.y
	31	33	7.2	649	12	BI261742	BI261742 602954071
c	32	33	7.2	826	28	BZ462870	BZ462870 BONBW34TR
c	33	32.8	7.2	521	14	CA703231	CA703231 wdk1c.pk0
	34	32.8	7.2	572	12	BI626794	BI626794 RH67848.5
	35	32.8	7.2	594	12	BJ335653	BJ335653 BJ335653
	36	32.8	7.2	613	14	CB935126	CB935126 ru03d09.y
	37	32.8	7.2	653	12	BJ336964	BJ336964 BJ336964
	38	32.8	7.2	1893	29	AY414592	AY414592 Pan trogl
	39	32.6	7.1	204	28	AZ577991	AZ577991 17d03 Sho
c	40	32.6	7.1	352	9	AA784257	AA784257 d3g05a1.f
c	41	32.6	7.1	356	9	AA785291	AA785291 g6c08a1.f
	42	32.6	7.1	778	29	CNS037J5	AL231386 Tetraodon
	43	32.6	7.1	813	28	BZ204853	BZ204853 CH230-390
	44	32.6	7.1	919	14	CK246169	CK246169 EST729806
	45	32.6	7.1	942	14	CK247920	CK247920 EST731557

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
ORIGIN

Query Match 52.1%; Score 237.4; DB 12; Length 680;
Best Local Similarity 70.0%; Pred. No. 6.4e-57;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60
              |||||||||||||||||| |||||| | |||||||||||| | || || || ||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120
              || || ||||| || |  |||||||||||||| ||     || ||||| |||| |  ||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              ||   | ||    | ||| |||||| |||||||||  || ||||  ||| ||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              |  |||||||||||| | |||    |     ||||||     ||     |  |     ||| ||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138
```

RESULT 2
AV430994/c
LOCUS AV430994 484 bp mRNA linear EST 23-AUG-2000
DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone PL027e10_r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
REFERENCE 1 (bases 1 to 484)
AUTHORS Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and Tabata,S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
PUBMED 10907854
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PL027e10_r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 10;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
Qy        321 CGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCGCGCTGGTTAA 380
              | | ||| || |   | |  | | ||  ||| |||||   ||| ||||| ||||||  |
Db        441 CCCAAAACACACAAGTGTCCCGGCCGCTCAAGACGGCCTCAATCACGCCTCGCTGGCCAG 382

Qy        381 TCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTTGGTTTTGGCAACAACGCCA 439
              | ||  ||   ||  |||| ||| ||||||     |  ||    |||  |  || ||
Db        381 GCCGAAAGCTGGTGTATCCAAAGTACTGGTGCAGGGTGCGGGAAAGGGCGTCGCCGGCA 323
```

RESULT 3
CA141136/c
LOCUS CA141136 664 bp mRNA linear EST 24-SEP-2003
DEFINITION SCJFRT2058D10.g RT2 Saccharum officinarum cDNA clone SCJFRT2058D10 5', mRNA sequence.
ACCESSION CA141136
VERSION CA141136.1 GI:35034821
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE 1 (bases 1 to 664)
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at http://www.bcccenter.fcav.unesp.br
Plate: 058 row: D column: 10
Seq primer: T7 Promoter Primer.
FEATURES Location/Qualifiers
source 1. .664
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCJFRT2058D10"
/lab_host="DH10B"
/clone_lib="RT2"
/note="Organ: Root tips(0.3cm-long) from adult plants; Vector: pSport1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from [Root tips(0.3cm-long) from adult plants]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library

construction can be obtained at
http://sucest.lad.ic.unicamp.br/public"
ORIGIN

Query Match 7.8%; Score 35.4; DB 13; Length 664;
Best Local Similarity 46.2%; Pred. No. 44;
Matches 117; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy         22 GCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGC 81
              || | ||      ||| || |   |    || |   | ||  | | ||  |||  | | |
Db        521 GCCTGCGAGTATGTCGGAGCTGAGGCACCTGGTTGCGTTGCAGACCTTGTACATGGCGAC 462

Qy         82 GGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATT 141
              |   |||      |   |  |||||| |      ||||| || |   |   |||||
Db        461 GTGCGCGCGGGGCGCCATCCCGGCGGCGACCCCCACGGCCTGGCCAAGAACCTGAGCGCC 402

Qy        142 TATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGAT 201
                ||  |||| |||| |  |  |||    | | |||    ||  ||| |       ||
Db        401 CTGCACGACGGCTCCGGTGGCCGTGCTGGCGGTGTGCGTGCCGTGGCCGGCGTCATCGAC 342

Qy        202 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAAT 261
              | ||          || ||    |  ||     | ||||| |||   | |  | ||| |
Db        341 CGGCGACGAGGAGTTGCTGCCGGCGGTGGGGACGGACATGAAGGAGCGAGCTCCGATGAG 282

Qy        262 AGTACTATTGAAC 274
              |  | ||| ||
Db        281 CTTGTTGTTGCAC 269
```

RESULT 4
BX251172
LOCUS BX251172 675 bp mRNA linear EST 24-FEB-2003
DEFINITION BX251172 Pinus pinaster differenciating xylem adult Pinus pinaster cDNA clone PP046C03, mRNA sequence.
ACCESSION BX251172
VERSION BX251172.1 GI:28513831
KEYWORDS EST.
SOURCE Pinus pinaster
ORGANISM Pinus pinaster
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 675)
AUTHORS Canton,F.R., Le Provost,G., Garcia,V., Barre,A., Frigerio,J.-M., Paiva,J., Fevereiro,P., Avila,C., Mouret,J.-F., Brach,J., de Daruvar,A., Canovas,F.M. and Plomion,C.
TITLE Transcriptome analysis of wood formation in maritime pine
JOURNAL Unpublished (2003)
COMMENT Contact: Frigerio JM
Genetique et Amelioration 69
INRA
route d'Arcachon 33612 Cestas CEDEX France
Email: Frigerio@pierroton.inra.fr
Email: Frigerio@pierroton.inra.fr.
FEATURES Location/Qualifiers
source 1. .675
/organism="Pinus pinaster"
/mol_type="mRNA"
/strain="ecotype: Corsican"
/db_xref="taxon:71647"
/clone="PP046C03"
/tissue_type="differenciating xylem"
/dev_stage="adult"
/clone_lib="Pinus pinaster differenciating xylem adult"
/note="Vector: Uni-Zap XR lambda (Stratagene); Site_1: Eco RI; Site_2: Xho I; A composite cDNA library was made with mRNA isolated from normal, compression, opposite, early and late wood of Maritime pine uni-directionally cloned into Uni-ZAP XR using the ZAP-cDNA Synthesis kit (Stratagene). pBluescript SK(-) plasmids were obtained by in vivo mass excision. The nucleotide sequence of the 5'end was obtained by automated sequencing with the T3 primer by GENOME EXPRESS, Meylan, France"
ORIGIN

Query Match 7.6%; Score 34.8; DB 13; Length 675;
Best Local Similarity 57.3%; Pred. No. 66;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        192 TAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGG 251
              ||||  |  |||| | | | |   | || | || ||||||   ||| |||||   | |||
Db        526 TAAAATTCCTCAGATAGATGCTATTATTTTGACACATGAACATGCAGATGCATTGCTGGG 585

Qy        252 TGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATG 301
              |  |||| ||  ||  | || ||   |  | ||    | ||   ||| ||
Db        586 TTTGGATGATGTTAGGAGTGTACAACCATATAACCCGTACAATGATATTG 635
```

RESULT 5
BI959328/c
LOCUS BI959328 699 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEn0019E09f Hordeum vulgare rachis EST library HVcDNA0015 (normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEn0019E09f, mRNA sequence.
ACCESSION BI959328
VERSION BI959328.1 GI:16310583
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 699)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex rachis cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 275
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence stop: 452.
FEATURES Location/Qualifiers
source 1. .699
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEn0019E09f"
/tissue_type="Rachis"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare rachis EST library HVcDNA0015 (normal)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and

sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 699;
Best Local Similarity 50.3%; Pred. No. 76;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy         37 GTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCAT 96
              ||| ||  ||||   ||||  | ||  ||||        || |||| |||| ||||     |
Db        202 GTACTTGGTGGCGTCGCTCCCGTTGAGGTCGACGTCGAAGGGGGAGGCGTCGGTGGCGTT 143

Qy         97 AACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTTCC 156
              || | |  ||||       ||||| |||  ||  |  | ||    | ||   ||
Db        142 AAGGTCCACGGCGCCGCCGACGGCCAGGAAGCAGAGAAGCGCCGCGAGCACGGCGAGGGA 83

Qy        157 GCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGTAAATATGATCAGC 205
              |   | |  || | |||   ||   | | ||   ||   | ||  | ||
Db         82 GGGGAGGGAGCCCATGCCGCGCGCTGTGGTGTGTGTGTGTGTGTGCGGC 34
```

RESULT 6
AV058630
LOCUS AV058630 296 bp mRNA linear EST 23-JUN-1999
DEFINITION AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA clone 1810054H02, mRNA sequence.
ACCESSION AV058630
VERSION AV058630.1 GI:5158377
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 296)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"
/clone_lib="Mus musculus pancreas C57BL/6J adult"

ORIGIN

Query Match 7.5%; Score 34.2; DB 9; Length 296;
Best Local Similarity 53.3%; Pred. No. 68;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        251 GTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCG 310
              |||  ||  | |||  | |  |  | ||  ||  ||| |  |||||  | ||  | |  |
Db         13 GTGGAGACCACAGTTTTTTCTAGGTCACCAAGGCTGGGTAAAGAAAAGAAGCATCAACTG 72

Qy        311 ACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAATAACGCCG 370
              ||||||  | | | |||||    ||    | ||    || | | |  || | ||| |  |
Db         73 TCCAGTGAAGCACAAAAAATGAAGAAGAAAATGATTACCCAAAAGTTGGAATTAAGGTAG 132

Qy        371 CGCTGGTTAATCAGA 385
              | ||   ||  |||
Db        133 AGATGCCAAAATAGA 147
```

RESULT 7
AZ959284/c
LOCUS AZ959284 669 bp DNA linear GSS 27-APR-2001
DEFINITION 2M0226G20R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0226G20 R, genomic survey sequence.
ACCESSION AZ959284
VERSION AZ959284.1 GI:13830511
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 669)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0226 row: G column: 20
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 669.
FEATURES Location/Qualifiers
source 1. .669
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0226G20"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 7.5%; Score 34.2; DB 28; Length 669;
Best Local Similarity 54.3%; Pred. No. 97;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              | |||  ||   || ||  || |  || |  ||  || |  |  || |||| ||||||
Db        235 CTGGAGGCATGCTTAAGGTTTCAGAAGCAATGTAACGTTCCCATTGTGCTTTCTCTGCTT 176

Qy        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240
              | | |  || |||||||| |     | |  || || ||  |  | |  | |||| |||
Db        175 CCTGCTTCAGGTCAATATGAGCTCTCAGCTGTCCCTGCTGCCATGCCTTTACTGGCCCAT 116

Qy        241 GCAGGCC 247
              || |
Db        115 CATGGAC 109
```

RESULT 8
BZ569281
LOCUS BZ569281 978 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_8328.y2 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_8328, genomic survey sequence.
ACCESSION BZ569281
VERSION BZ569281.1 GI:27203648
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 978)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .978
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_8328"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN

Query Match 7.5%; Score 34.2; DB 28; Length 978;
Best Local Similarity 53.3%; Pred. No. 1.1e+02;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        213 CCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGA 272
              | || | | |||||||   || | | |||  ||| ||   |  ||| | || | ||||||
Db        519 CGGTATCGATACCCATTCGATTGAAAATGATGGCGAGTTCGTTGATCAGAGCAATATTGA 578

Qy        273 ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTC 332
              | | | |  | ||||    |   || || | ||  | | |  || |||     |
Db        579 GATCTCGCTGCGTATTTTCGATCACCTTGGCAGCCTCAGCGACTTTAATGCTGCTCGCCT 638

Qy        333 CGATATTACTGTCGG 347
              |  | ||| | | |
Db        639 TGTGAGTACCGGCAG 653
```

RESULT 9
BG671547
LOCUS BG671547 1036 bp mRNA linear EST 30-APR-2001
DEFINITION DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5', mRNA sequence.
ACCESSION BG671547
VERSION BG671547.1 GI:13893646
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1036)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
TITLE Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
MEDLINE 22056133
PUBMED 12060780
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR PRimers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1036
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRNBUB05"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/clone_lib="Rat DRG Library"

ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 1036;
Best Local Similarity 49.2%; Pred. No. 1.2e+02;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy         11 TAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTC 70
              | ||||   ||  | | |||    |||  || | | || |||| |    ||||| |
Db        168 TGAAAGATACATGAGTGACAGTGTCCGTGATTGCAGACATTGCTGTATGATGCGTCCTGT 227

Qy         71 CACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAA 130
              |   | | | |  ||  |||||  ||  ||  | |   || | |  |  ||    |
Db        228 CTGTAGGAGTCCTCGCTGGTAACGATGCCGAGAGTGCACATCGATATCGGTGGGCGGTTA 287
```

```
Qy        131 CGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCC 190
              | |       |  |  ||   | |    | ||| |||   ||| ||||  ||| || ||
Db        288 CATCTGAAGTGGAAGAGGTGGCTCTTTCGAACTCTGACTTTGATCTGTCAAGGGACGCTT 347

Qy        191 GTA 193
              |||
Db        348 GTA 350


RESULT 10
BZ565783/c
LOCUS       BZ565783                1137 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_5750.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_5750, genomic survey sequence.
ACCESSION   BZ565783
VERSION     BZ565783.1  GI:27194113
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1137)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_5750"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             7.5%;  Score 34.2;  DB 28;  Length 1137;
  Best Local Similarity   53.3%;  Pred. No. 1.2e+02;
  Matches   72;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        213 CCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGTGCGGATAATAGTACTATTGA 272
              | || | | |||||||   || | | |||  ||| ||   |  ||| | || | ||||||
Db        272 CGGTATCGATACCCATTCGATTGAAAATGATGGCGAGTTCGTTGATCAGAGCAATATTGA 213

Qy        273 ACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAAAACTC 332
              | | | |  ||||    |  || || | ||  | | |  || |||     |
Db        212 GATCTCGCTGCGTATTTTCGATCACCTTGGCAGCCTCAGCGACTTTAATGCTGCTCGCCT 153

Qy        333 CGATATTACTGTCGG 347
              |  | ||| | | |
Db        152 TGTGAGTACCGGCAG 138


RESULT 11
BB581987/c
LOCUS       BB581987                234 bp    mRNA    linear   EST 30-NOV-2000
DEFINITION  BB581987 RIKEN full-length enriched, adult male colon Mus musculus
            cDNA clone 9030005O05 5', mRNA sequence.
ACCESSION   BB581987
VERSION     BB581987.1  GI:11478531
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
            Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
            Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
            Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
            Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
            Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9030005O05"
                     /sex="male"
                     /tissue_type="colon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male colon"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 10;  Length 234;
```

```
  Best Local Similarity   51.3%;  Pred. No. 70;
  Matches   79;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         35 TCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACAATGGGGCGGCGGCGGTAATC 94
              | ||  ||  | |     | |   || ||  || |     ||    || | || || ||    |
Db        200 TAGTTTTTGGTAGGTCGGTTTATGCGGGTTTCTTCTTCCACCAGGCCCGCCGCCGTTTCC 141

Qy         95 ATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTGAGCATTTATCAGTACGGTT 154
                |||| |||  ||  ||   ||        |   |  || | ||||||| ||  ||  ||
Db        140 CGAACGCCGGGTGCTTTATAGCCTTGTTCAAAATATGGTAGCGCATTTACCACCACCTTT 81

Qy        155 CCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGC 188
              ||| | ||   || | ||| |  ||  |||  ||
Db         80 CCGGTTACCAAGCCCGTGCCCGCCATTGCGCGGC 47


RESULT 12
CA711477/c
LOCUS       CA711477                500 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wdk2c.pk014.o3 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.o3 5'
            end, mRNA sequence.
ACCESSION   CA711477
VERSION     CA711477.1  GI:25433270
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone="wdk2c.pk014.o3"
                     /tissue_type="kernel"
                     /clone_lib="wdk2c"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
                     days after anthesis."
ORIGIN

  Query Match             7.5%;  Score 34;  DB 14;  Length 500;
  Best Local Similarity   54.5%;  Pred. No. 97;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  |  ||| |||||||   ||   |  | ||| ||       | | ||  | ||| |
Db        274 AGNTGNTGCAGTCGCAGCCNCCGGTCTCCCCGGCCGTNTCGAACCCGCCGGAGCTCCCCN 215

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGAC 126
                 ||| ||||| ||  || |  | |||||||||     |    ||||   ||
Db        214 TGTGGGAGGCGGTGGGCATGACCATGGCGGCGGCGCCGGCGGTGGCCNCCAC 163


RESULT 13
AZ180696
LOCUS       AZ180696                500 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_0181_A1_C06_SP6E Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library Strongylocentrotus purpuratus
            genomic clone Plate=181 Col=11 Row=E, genomic survey sequence.
ACCESSION   AZ180696
VERSION     AZ180696.1  GI:8353071
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
            Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H.
            and Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
   PUBMED   10920195
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 181  row: E  column: 11
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 500.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7668"
                     /clone="Plate=181 Col=11 Row=E"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
ORIGIN

  Query Match             7.5%;  Score 34;  DB 28;  Length 500;
  Best Local Similarity   50.3%;  Pred. No. 97;
  Matches   82;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        133 TTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGCGATGCCCGT 192
              ||   || | ||    | | |  || | | | ||    || |    |    ||||   |
Db          8 TTNTCCAATAATATCCATGATAACGATGATGATGATGATGATGATGATGATTGA 67

Qy        193 AAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCCAGGGT 252
                | ||||| |    | |      || || |    |||||   |   |||  |   | | |
Db         68 GACGATGATGACGATGATGATGATGATGATGATGATGAATACGATGATGATGATGATGAT 127

Qy        253 GCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAA 295
              |  ||| ||| || || |||| | | | |  ||| || |||||
Db        128 GATGATGATAATAATAATGAAATTAATAATGATGATTACAGAA 170


RESULT 14
AA746477/c
LOCUS       AA746477                365 bp    mRNA    linear   EST 27-JAN-1998
DEFINITION  nw62e02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
            similar to TR:O00555 O00555 NEURONAL CALCIUM CHANNEL ALPHA 1A
            SUBUNIT ISOFORM A-1. ;contains element TAR1 repetitive element ;,
            mRNA sequence.
ACCESSION   AA746477
VERSION     AA746477.1  GI:2786463
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 365)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 661    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1251194"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 365;
  Best Local Similarity   56.9%;  Pred. No. 97;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         15 AGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCTGGCGTCGTTCCACA 74
              ||  ||||||   ||||||   | ||   | ||| |||  | ||| |||   |  ||||
Db        119 AGCAGCAGCAGCAGCAGCAGCAGCAGGCGGTGGCCAGGCCGGCCGGGCGCCACC 60

Qy         75 ATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 123
              |  || |  ||| ||||  ||   | ||   | | |   | | |||| |
Db         59 AGCGGCCCTCGGAGGTACCCAGGCCCGCACGGCCGACGCTACTGGCCGG 11


RESULT 15
CD374421/c
LOCUS       CD374421                 558 bp    mRNA    linear   EST 19-AUG-2003
DEFINITION  TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)
            Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION   CD374421
VERSION     CD374421.1  GI:31229939
KEYWORDS    EST.
SOURCE      Wuchereria bancrofti
  ORGANISM  Wuchereria bancrofti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Wuchereria.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
  TITLE     Toward the characterization of the Wuchereria bancrofti
            microfilarial transcriptome with comparisons to those of Brugia
            malayi and Onchocerca volvulus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Thomas B. Nutman
            Laboratory of Parasitic Diseases
            NIH
            Building 4 Room 126, Bethesda, MD 20892-0425, USA
            Email: tnutman@niaid.nih.gov
            Seq primer: pBluescript T3.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Wuchereria bancrofti"
                     /mol_type="mRNA"
                     /db_xref="taxon:6293"
                     /clone="TNWbmfC2H3"
                     /dev_stage="microfilaria"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="Wuchereria bancrofti microfilaria cDNA
                     (SAW95SjL-WbMf)"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 85,000 microfilariae
                     isolated from the blood of an infected individual from
                     Guyana, South America and converted to double-stranded
                     cDNA using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNA pol I. The library has 2.9 x 10E5
                     independent recombinants and the average insert size is
                     ~1kb. The library was constructed by Sandra J. Laney. The
                     library is available from Dr.S.A.Williams, email:
                     genome@smith.edu."
ORIGIN

  Query Match             7.4%;  Score 33.8;  DB 14;  Length 558;
  Best Local Similarity   60.7%;  Pred. No. 1.2e+02;
  Matches   65;  Conservative    0;  Mismatches   40;  Indels    2;  Gaps    1;

Qy         76 TGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCGGACTCAACGTTG 135
              ||| |||| |||||||| |     || ||||| |||| |    | ||  | |
Db        161 TGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCCCAAATTCTANGCT 102

Qy        136 AGCATTTATC--AGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180
              |||||| |  ||  ||||  ||  | | |||| |||||||| ||
Db        101 GACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55


Search completed: March 18, 2004, 20:11:33
Job time : 2035.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		456	100.0	2067	1	SEU43280	U43280 Salmonella
2		452.8	99.3	5103	1	STAJ2301	AJ002301 Salmonell
3		452.8	99.3	22411	1	AE008749	AE008749 Salmonell
4		451.2	98.9	254050	1	AL627269	AL627269 Salmonell
5	c	451.2	98.9	301983	1	AE016840	AE016840 Salmonell
6		382.8	83.9	1048	1	STAGFBA	AJ000514 Salmonell
7		348.8	76.5	2889	1	CSP515700	AJ515700 Citrobact
8		334.4	73.3	10370	1	AE015131	AE015131 Shigella
9		332.8	73.0	4680	1	ECCSGABDG	X90754 E.coli csgG
10		332.8	73.0	10346	1	AE000205	AE000205 Escherich
11		332.8	73.0	15047	1	D90741	D90741 Escherichia
12		332.8	73.0	292504	1	AE016981	AE016981 Shigella
13		332.8	73.0	306358	1	AE016759	AE016759 Escherich
14		331.2	72.6	1711	1	AF275733	AF275733 Escherich
15		331.2	72.6	10190	1	AE005315	AE005315 Escherich
16		331.2	72.6	327773	1	AP002554	AP002554 Escherich
17		313.4	68.7	2920	1	CFR515701	AJ515701 Citrobact
18		271	59.4	2883	1	ESA515702	AJ515702 Enterobac
19		184	40.4	1212	1	ECO131756	AJ131756 Escherich
20	c	47.8	10.5	10709	1	AE015532	AE015532 Shewanell
21		47.4	10.4	301214	1	AE016786	AE016786 Pseudomon
22		44.2	9.7	11160	1	AE008209	AE008209 Agrobacte
23	c	44.2	9.7	11476	1	AE009405	AE009405 Agrobacte
24		41.4	9.1	42602	2	AC116971	AC116971 Trypanoso
25	c	41	9.0	7218	6	I66494	I66494 Sequence 14
26	c	40.6	8.9	62381	5	BX640594	BX640594 Zebrafish
27		40.4	8.9	168457	2	AC111425	AC111425 Rattus no
28	c	40.4	8.9	178934	2	AC127966	AC127966 Rattus no
29		40.2	8.8	193553	10	AC124196	AC124196 Mus muscu
30	c	39.8	8.7	295500	1	AP005954	AP005954 Bradyrhiz
31	c	38.2	8.4	298406	3	CEY75B8A	AL033514 Caenorhab
32	c	37.8	8.3	174732	2	AL136161	AL136161 Homo sapi
33	c	37.8	8.3	191567	9	AL138702	AL138702 Human DNA
34	c	37.6	8.2	2000	6	AX655393	AX655393 Sequence
35		37.4	8.2	120937	10	AL591911	AL591911 Mouse DNA
36	c	37.4	8.2	225016	2	AC063967	AC063967 Mus muscu
37		37.2	8.2	42310	3	CBRG01B5	AC084425 Caenorhab
38	c	37	8.1	164504	2	AC123240	AC123240 Rattus no
39		36.8	8.1	252420	3	AE014841	AE014841 Plasmodiu
40	c	36.6	8.0	92378	5	BX537249	BX537249 Zebrafish
41	c	36.6	8.0	110000	2	AC095863_06	Continuation (7 of
42		36.6	8.0	196383	2	AC113159	AC113159 Rattus no
43		36.6	8.0	230058	2	AC108229	AC108229 Rattus no
44		36.6	8.0	254497	2	AC119520	AC119520 Rattus no
45		36.6	8.0	323930	1	AP003194	AP003194 Clostridi

ALIGNMENTS

RESULT 1
SEU43280
LOCUS SEU43280 2067 bp DNA linear BCT 14-FEB-1996
DEFINITION Salmonella enteritidis agfBAC operon: fimbrin-like protein precursor (agfB), thin aggregative fimbriae precursor (agfA), and AgfC (agfC) genes, complete cds.
ACCESSION U43280
VERSION U43280.1 GI:1184712
KEYWORDS .
SOURCE Salmonella enteritidis
ORGANISM Salmonella enteritidis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 2067)
AUTHORS Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,

```
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN

  Query Match           100.0%;  Score 456;  DB 1;  Length 2067;
  Best Local Similarity  100.0%;  Pred. No. 3.8e-123;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        696 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 755

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 815

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        816 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 875

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        876 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 935

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        936 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 995

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        996 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 1055

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1056 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 1115

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1116 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 1151


RESULT 2
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
     gene            3344. .3376
                     /gene="csgBA"
     -35_signal      3344. .3349
                     /gene="csgBA"
     -10_signal      3371. .3376
                     /gene="csgBA"
     gene            3469. .3924
                     /gene="csgB"
     CDS             3469. .3924
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="CAA05316.1"
                     /db_xref="GI:2739238"
                     /db_xref="GOA:P55226"
                     /db_xref="SWISS-PROT:P55226"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     3469. .3531
                     /gene="csgB"
     mat_peptide     3532. .3921
                     /gene="csgB"
                     /product="CsgB protein"
                     /function="nucleator"
     gene            3966. .4421
                     /gene="csgA"
     CDS             3966. .4421
                     /gene="csgA"
                     /function="fiber subunit"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="CAA05317.1"
                     /db_xref="GI:2739239"
                     /db_xref="GOA:P55225"
                     /db_xref="SWISS-PROT:P55225"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     3966. .4025
                     /gene="csgA"
     mat_peptide     4026. .4418
                     /gene="csgA"
                     /product="CsgA protein"
                     /function="fiber subunit"
     gene            4483. .4809
                     /gene="csgC"
     CDS             4483. .4809
                     /gene="csgC"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgC protein"
                     /protein_id="CAA05318.1"
```

```
                     /db_xref="GI:2739240"
                     /db_xref="SWISS-PROT:P55227"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
     sig_peptide     4483. .4533
                     /gene="csgC"
     mat_peptide     4534. .4806
                     /gene="csgC"
                     /product="CsgC"
     gene            4940. .5103
                     /gene="ORF103"
     CDS             4940. .>5103
                     /gene="ORF103"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA05319.1"
                     /db_xref="GI:2739241"
                     /db_xref="SWISS-PROT:O54295"
                     /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                     CDISTHAENI"
ORIGIN

  Query Match             99.3%;  Score 452.8;  DB 1;  Length 5103;
  Best Local Similarity   99.6%;  Pred. No. 3.4e-122;
  Matches  454;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3469 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 3528

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       3529 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 3588

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3589 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 3648

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3649 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 3708

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3709 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 3768

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       3769 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 3828

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3829 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 3888

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       3889 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 3924


RESULT 3
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES                Location/Qualifiers
     source             1. .22411
                        /organism="Salmonella typhimurium LT2"
                        /mol_type="genomic DNA"
                        /strain="LT2; SGSC 1412; ATCC 700720"
                        /db_xref="ATCC:700720"
                        /db_xref="taxon:99287"
                        /note="LT2"
     gene               434. .1308
                        /gene="phoH"
                        /note="synonym: STM1126"
     RBS                434. .439
                        /gene="phoH"
                        /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS                454. .1308
                        /gene="phoH"
                        /note="similar to E. coli PhoB-dependent, ATP-binding pho
                        regulon component; may be helicase; induced by P
                        starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                        aa), 92% identity in aa 71 - 354"
                        /codon_start=1
                        /transl_table=11
                        /product="PhoB-dependent ATP-binding pho regulon
                        component"
                        /protein_id="AAL20057.1"
                        /db_xref="GI:16419642"
                        /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                        RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                        GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                        YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                        LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                        QRTLHAYS"
     gene               complement(1414. .2303)
```

```
                /gene="STM1127"
     CDS        complement(1414. .2295)
                /gene="STM1127"
                /note="similar to E. coli orf, hypothetical protein
                (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                identity in aa 15 - 264"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /protein_id="AAL20058.1"
                /db_xref="GI:16419643"
                /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                QARESKLRTMNALDMTK"
     RBS        complement(2298. .2303)
                /gene="STM1127"
                /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene       complement(2580. .4088)
                /gene="STM1128"
     CDS        complement(2580. .4076)
                /gene="STM1128"
                /note="similar to E. coli putative cotransporter
                (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                identity in aa 7 - 478"
                /codon_start=1
                /transl_table=11
                /product="putative sodium/glucose cotransporter"
                /protein_id="AAL20059.1"
                /db_xref="GI:16419644"
                /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITPLTVLALRPFIAIDPVILVL
                LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                DKFFPATQFHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS        complement(4083. .4088)
                /gene="STM1128"
                /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene       complement(4413. .5099)
                /gene="STM1129"
     CDS        complement(4413. .5093)
                /gene="STM1129"
                /note="similar to E. coli putative enzyme (AAC76255.1);
                Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                226"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL20060.1"
                /db_xref="GI:16419645"
                /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                WFSHAVKR"
     RBS        complement(5094. .5099)
                /gene="STM1129"
                /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene       5599. .6759
                /gene="STM1130"
     CDS        5599. .6759
                /gene="STM1130"
                /note="similar to E. coli orf, hypothetical protein
                (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                identity in aa 40 - 403"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL20061.1"
                /db_xref="GI:16419646"
                /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene       6793. .7497
                /gene="STM1131"
     RBS        6793. .6798
                /gene="STM1131"
                /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
     CDS        6805. .7497
                /gene="STM1131"
                /note="similar to E. coli orf, hypothetical protein
                (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                identity in aa 4 - 239"
                /codon_start=1
                /transl_table=11
                /product="putative outer membrane protein"
                /protein_id="AAL20062.1"
                /db_xref="GI:16419647"
                /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                QTRYRVGLQYHF"
     gene       7768. .9060
                /gene="STM1132"
     RBS        7768. .7773
                /gene="STM1132"
                /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS        7780. .9060
                /gene="STM1132"
                /note="similar to E. coli putative transport protein
                (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                identity in aa 25 - 418"
                /codon_start=1
                /transl_table=11
                /product="putative sugar transport protein"

  Query Match             99.3%;  Score 452.8;  DB 1;  Length 22411;
  Best Local Similarity   99.6%;  Pred. No. 3.6e-122;
  Matches  454;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17272 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 17331

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      17332 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 17391

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17392 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 17451

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17452 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 17511

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17512 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 17571

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      17572 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 17631

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17632 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 17691

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
                ||||||||||||||||||||||||||||||||||||
Db      17692 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 17727


RESULT 4
AL627269
LOCUS       AL627269               254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAARAKIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
     misc_feature    complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
     gene            5924. .8032
                     /gene="STY1082"
     CDS             5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
     misc_feature    5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
     misc_feature    6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
     gene            8045. .9952
                     /gene="STY1083"
     CDS             8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap

  Query Match           98.9%;  Score 451.2;  DB 1;  Length 254050;
  Best Local Similarity  99.3%;  Pred. No. 1.2e-121;
  Matches  453;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88409 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 88468

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      88469 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 88528

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88529 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 88588

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      88589 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGAAAATAATCGGGCG 88648

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88649 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 88708

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      88709 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 88768

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88769 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 88828

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      88829 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 88864


RESULT 5
AE016840/c
LOCUS       AE016840              301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
```

/transl_table=11
/product="transposase for insertion sequence element IS200"
/protein_id="AAO69359.1"
/db_xref="GI:29137798"
/translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"

gene complement(844. .1797)
/gene="flgL"
/locus_tag="t1736"

CDS complement(844. .1797)
/gene="flgL"
/locus_tag="t1736"
/note="corresponds to STY1223 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook-associated protein 3"
/protein_id="AAO69360.1"
/db_xref="GI:29137799"
/translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"

gene complement(1812. .3473)
/gene="flgK"
/locus_tag="t1737"

CDS complement(1812. .3473)
/gene="flgK"
/locus_tag="t1737"
/note="corresponds to STY1222 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook-associated protein 1"
/protein_id="AAO69361.1"
/db_xref="GI:29137800"
/translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"

gene complement(3538. .4488)
/gene="flgJ"
/locus_tag="t1738"

CDS complement(3538. .4488)
/gene="flgJ"
/locus_tag="t1738"
/note="corresponds to STY1221 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar protein FlgJ"
/protein_id="AAO69362.1"
/db_xref="GI:29137801"
/translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"

gene complement(4488. .5591)
/gene="flgI"
/locus_tag="t1739"

CDS complement(4488. .5591)
/gene="flgI"
/locus_tag="t1739"
/note="corresponds to STY1220 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar P-ring protein precursor"
/protein_id="AAO69363.1"
/db_xref="GI:29137802"
/translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGGSLQSVRSSANLNSVV RALNALGATPMDLMSILQSMQSAGCLRAKLEII"

gene complement(5597. .6295)
/gene="flgH"
/locus_tag="t1740"

CDS complement(5597. .6295)
/gene="flgH"
/locus_tag="t1740"
/note="corresponds to STY1219 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AAO69364.1"
/db_xref="GI:29137803"
/translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN MGWLQRFFLNLSPM"

gene complement(6350. .7132)
/gene="flgG"
/locus_tag="t1741"

CDS complement(6350. .7132)
/gene="flgG"
/locus_tag="t1741"
/note="distal rod protein; corresponds to STY1218 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein FlgG"
/protein_id="AAO69365.1"
/db_xref="GI:29137804"
/translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"

gene complement(7146. .7901)
/gene="flgF"
/locus_tag="t1742"

CDS complement(7146. .7901)
/gene="flgF"
/locus_tag="t1742"
/note="corresponds to STY1217 from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein FlgF"
/protein_id="AAO69366.1"
/db_xref="GI:29137805"
/translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT DMIANARRFEMQMKVITSVDENEGRANQLLSMS"

gene complement(7922. .9133)
/gene="flgE"
/locus_tag="t1743"

```
     CDS             complement(7922. .9133)
                     /gene="flgE"
                     /locus_tag="t1743"
                     /note="corresponds to STY1216 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook protein FlgE"
                     /protein_id="AAO69367.1"
                     /db_xref="GI:29137806"
                     /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                     GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL

  Query Match             98.9%;  Score 451.2;  DB 1;  Length 301983;
  Best Local Similarity   99.3%;  Pred. No. 1.2e-121;
  Matches  453;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37807 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 37748

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      37747 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 37688

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37687 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 37628

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      37627 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGAAAATAATCGGGCG 37568

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37567 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 37508

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      37507 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 37448

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37447 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 37388

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      37387 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 37352


RESULT 6
STAGFBA
LOCUS       STAGFBA                 1048 bp    DNA     linear   BCT 26-JAN-1998
DEFINITION  Salmonella typhimurium agfB and agfA genes.
ACCESSION   AJ000514
VERSION     AJ000514.1  GI:2275119
KEYWORDS    agfA gene; agfB gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
            Normark,S.J. and Rhen,M.
  TITLE     Expression of thin aggregative fimbriae promotes interaction of
            Salmonella typhimurium SR-11 with mouse small intestinal epithelial
            cells
  JOURNAL   Infect. Immun. 65 (12), 5320-5325 (1997)
  MEDLINE   98053981
   PUBMED   9393832
REFERENCE   2  (bases 1 to 1048)
  AUTHORS   Sukupolvi,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
            University of Turku, Kiinanmyllynkatu, 20520, FINLAND
FEATURES             Location/Qualifiers
     source          1. .1048
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:602"
     gene            14. .553
                     /gene="agfB"
     CDS             14. .553
                     /gene="agfB"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB protein"
                     /protein_id="CAA04150.1"
                     /db_xref="GI:2275120"
                     /db_xref="SPTREMBL:O33801"
                     /translation="MKNKLLFMMLTILGEPIATATNYDLARSEYNFAVNELSKSSFNQ
                     AAIIGQVGTDNSARVRHEGSKLLSVISQERGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAISAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIQANITQYGTQ
                     KTAVVVQKQSHMAIRVTQR"
     gene            593. .1048
                     /gene="agfA"
     CDS             593. .>1048
                     /gene="agfA"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA protein"
                     /protein_id="CAA04151.1"
                     /db_xref="GI:2275121"
                     /db_xref="SPTREMBL:O33802"
                     /translation="MKLLKVAAFAAIVVSGSAVAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNAPANQYN"
ORIGIN

  Query Match             83.9%;  Score 382.8;  DB 1;  Length 1048;
  Best Local Similarity   94.6%;  Pred. No. 1.3e-101;
  Matches  435;  Conservative    0;  Mismatches    7;  Indels   18;  Gaps    3;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||||||||||||||||| |||   ||||||||||
Db         14 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGAGCC---GATTGCAACC 70

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         71 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 130

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 190

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGA---AGGAGGAAATAATCGG 237
              |||||| |||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        191 CGCCATGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAACGAGGAGGAAATAATCGG 250

Qy        238 GCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCC 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 GCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCC 310

Qy        298 AACGATGCCAGTATATCGCAAAGCGCTTACGGTAA------------TAGTGCAGCTATT 345
              |||||||||||||||||||||||||||||||||||            |||||| ||||||
Db        311 AACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTAGTGCGGCTATT 370

Qy        346 ATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCA 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 ATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCA 430

Qy        406 GTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCA 445
              |||||||||||||||||||||||||||||||||||  | ||
Db        431 GTTGTAGTGCAGAAACAGTCGCATATGGCTATTCAGGCCA 470
```

```
RESULT 7
CSP515700
LOCUS       CSP515700               2889 bp    DNA     linear   BCT 24-JUN-2003
DEFINITION  Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION   AJ515700
VERSION     AJ515700.1  GI:31790491
KEYWORDS    csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
            component of curlin monomers; regulatory protein.
SOURCE      Citrobacter sp. Fec2
  ORGANISM  Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
  AUTHORS   Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
  TITLE     Production of Cellulose and Curli Fimbriae by Members of the Family
            Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
  JOURNAL   Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Citrobacter sp. Fec2"
                     /mol_type="genomic DNA"
                     /isolate="Fec2"
                     /isolation_source="faeces"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:213763"
                     /country="Germany"
     gene            complement(212. .862)
                     /gene="csgD"
     CDS             complement(212. .862)
                     /gene="csgD"
                     /function="necessary for transcription of the csgBA
                     operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAD56670.1"
                     /db_xref="GI:31790492"
                     /translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
                     RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
                     HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
                     KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            1623. .2078
                     /gene="csgB"
     CDS             1623. .2078
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAD56671.1"
                     /db_xref="GI:31790493"
                     /translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
                     QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
                     SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
     gene            2119. .2571
                     /gene="csgA"
     CDS             2119. .2571
                     /gene="csgA"
                     /function="curli subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin-csgA protein"
                     /protein_id="CAD56672.1"
                     /db_xref="GI:31790494"
                     /translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
                     SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
                     DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN
```

```
  Query Match           76.5%;  Score 348.8;  DB 1;  Length 2889;
  Best Local Similarity  85.3%;  Pred. No. 1.4e-91;
  Matches  389;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||| ||| ||||||||||||||| |||||  ||
Db       1623 ATGAAAAACAAATTGTTATTTATGATGTTAACAGTACTGGGTGCGCCTGGAATTGCTTCC 1682

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              || |||| ||||||| | || | ||| |||||||| |||||||||||||||||||| |||
Db       1683 GCAACAAGTTATGATTTAGCGCATTCCGAATATAACTTTGCGGTAAATGAATTAAGTAAG 1742

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| |||||||||||||||| |||||| |||||||||   |    |
Db       1743 TCTTCATTTAATCAGGCAGCCATTATTGGTCAAGTGGGCACGAATAATAGTGCAAAGATG 1802

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||| || || ||||| ||||| ||  |||||||||||||| |||| ||| |||||
Db       1803 CGCCAGGAGGGCTCTAAACTTTTGTCGGTAGTTTCACAAGAAGGTGGAAGTAACCGGGCC 1862

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||| |||||  ||||   |||||| ||||||||||||||  ||  ||||  || ||||
Db       1863 AAAGTTGACCAATCAGGAGCTTATAATTTTGCGTATATTGCTCAGTCGGGTCATTCCAAT 1922

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||   |||||||||| |||    |||||||||||||||||||
Db       1923 GATGCCAGTATATCGCAAAGTAATTACGGTAATACTGCGATGATTATCCAGAAAGGTTCT 1982

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              |||||||| || |||||||| ||||| |||||||| |||||||||||||||||||||| |
Db       1983 GGAAATAAAGCAAATATTACTCAGTATGGTACGCAAAAAACAGCAGTTGTAGTGCAGAGA 2042

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| ||||| |||||||| |  |||||||||
Db       2043 CAGTCGCAAATGGCAATTCGCGTTATTCAACGCTAA 2078
```

```
RESULT 8
AE015131
LOCUS       AE015131               10370 bp    DNA     linear   BCT 18-OCT-2002
DEFINITION  Shigella flexneri 2a str. 301 section 94 of 412 of the complete
            genome.
ACCESSION   AE015131 AE005674
VERSION     AE015131.1  GI:24051313
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 301
  ORGANISM  Shigella flexneri 2a str. 301
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 10370)
  AUTHORS   Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
            Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
            Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
            Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
            Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
  TITLE     Genome sequence of Shigella flexneri 2a: insights into
            pathogenicity through comparison with genomes of Escherichia coli
            K12 and O157
  JOURNAL   Nucleic Acids Res. 30 (20), 4432-4441 (2002)
   PUBMED   12384590
REFERENCE   2  (bases 1 to 10370)
  AUTHORS   Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B.,
            Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y.,
            Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y.,
            Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
            of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.
            China
FEATURES             Location/Qualifiers
     source          1. .10370
                     /organism="Shigella flexneri 2a str. 301"
```

/mol_type="genomic DNA"
/strain="301"
/serotype="2a"
/db_xref="taxon:198214"
tRNA complement(127. .214)
/product="tRNA-Ser"
/note="anticodon: GGA"
gene 410. .1387
/gene="ycdW"
/locus_tag="SF1027"
/pseudo
CDS 410. .1387
/gene="ycdW"
/locus_tag="SF1027"
/note="similar to Escherichia coli K12 putative dehydrogenase gi: 1787270 (326 aa). BLAST with identity of 99% in 326 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11
gene 1442. .2179
/gene="ycdX"
/locus_tag="SF1028"
CDS 1442. .2179
/gene="ycdX"
/locus_tag="SF1028"
/note="Residues 1 to 245 of 245 are 99 pct identical to residues 1 to 245 of a 245 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287168.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42652.1"
/db_xref="GI:24051314"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE PVFAPHDKATNTQAMISTIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS SFLHSRKGSEDNCRAVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPLER ILNVSPRRLLNFLESRGMAPIAEFADL"
gene 2203. .2757
/gene="ycdY"
/locus_tag="SF1029"
CDS 2203. .2757
/gene="ycdY"
/locus_tag="SF1029"
/note="Residues 1 to 184 of 184 are 99 pct identical to residues 1 to 184 of a 184 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287169.1"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase component"
/protein_id="AAN42653.1"
/db_xref="GI:24051315"
/translation="MNEPSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEDATEAEVRAFLSERGMLL ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW RTMAPLTRDAISAMWDELEEDSEE"
gene 2811. .3350
/gene="ycdZ"
/locus_tag="SF1030"
CDS 2811. .3350
/gene="ycdZ"
/locus_tag="SF1030"
/note="Residues 1 to 179 of 179 are 98 pct identical to residues 1 to 179 of a 179 aa protein from Escherichia coli K12 ref: NP_415554.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42654.1"
/db_xref="GI:24051316"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLVFGYAMKNSGLW LAARSAKTAHREQEIKNKA"
gene complement(3414. .4189)
/gene="csgG"
/locus_tag="SF1031"
/pseudo
CDS complement(3414. .4189)
/gene="csgG"
/locus_tag="SF1031"
/note="similar to Escherichia coli K12 curli production assembly-transport component, 2nd curli operon gi: 1787274 (278 aa). BLAST with identity of 97% in 257 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11
gene complement(4454. .5662)
/locus_tag="SF1032"
CDS complement(4454. .5662)
/locus_tag="SF1032"
/note="Residues 1 to 402 of 402 are 100 pct identical to residues 1 to 402 of a 402 aa protein from Plasmid R100 ref: NP_052925.1"
/codon_start=1
/transl_table=11
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translation="MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTLTLTEL GRNLPTKARTKHNIKRIDRLLGNRHLHKERLAVYRWHASFICSGNTMPIVLVDWSDIR EQKRLMVLRASVALHGRSVTLYEKAFPLSEQCSKKAHDQFLADLASILPSNTTPLIVS DAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAENWKPISNLHDMSSSHSKTLGYKR LTKSNPISCQILLYKSRSKGRKNQRSTRTHCHHPSPKIYSASAKEPWILATNLPVEIR TPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWL AGVHAQKQGWDKHFQANTVRNRNVLSTVRLGMEVLRHSGYTITREDSLVAATLLTQNL FTHGYVLGKL"
gene complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"
CDS complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"
/note="Residues 1 to 129 of 129 are 98 pct identical to residues 1 to 129 of a 129 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287173.1"
/codon_start=1
/transl_table=11
/product="curli production assembly-transport component, 2nd curli operon"
/protein_id="AAN42656.1"
/db_xref="GI:24051318"
/translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDSEKTVVFAL IQTEEALNRRQINQALLSTDDLAHDEF"
gene complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"
CDS complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"
/note="Residues 11 to 226 of 226 are 100 pct identical to residues 1 to 216 of a 216 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=11
/product="putative 2-component transcriptional regulator for 2nd curli operon"
/protein_id="AAN42657.1"
/db_xref="GI:24051319"
/translation="MRSIKKWGFIMFNEVHSIHGHTLLLITKPSLQATALLQHLKQSL

```
                AITGKLHNIQRSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPED
                YPYRDIENWPHINGVFYAMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNS
                TESALLTHREKEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVS
                WANDNLRR"
     gene       7766. .8248
                /gene="csgB"
                /locus_tag="SF1035"
     CDS        7766. .8248
                /gene="csgB"
                /locus_tag="SF1035"
                /note="Residues 10 to 160 of 160 are 99 pct identical to
                residues 1 to 151 of a 151 aa protein from Escherichia
                coli O157:H7 EDL933 ref: NP_287175.1"
                /codon_start=1
                /transl_table=11
                /product="minor curlin subunit precursor, similar ro CsgA"
                /protein_id="AAN42658.1"
                /db_xref="GI:24051320"
                /translation="MYDQVQGDNMKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAV
                NELSKSSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYI
                DQAGSANDASISQGAYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR

  Query Match             73.3%;  Score 334.4;  DB 1;  Length 10370;
  Best Local Similarity   83.3%;  Pred. No. 2.6e-87;
  Matches  380;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||
Db       7793 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 7852

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || ||||||||||||||| || |||
Db       7853 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 7912

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db       7913 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 7972

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  |||||||||
Db       7973 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAATCGGGCA 8032

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db       8033 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 8092

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db       8093 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 8152

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| |||||||||||||||| |
Db       8153 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAGTTGTAGTGCAGAGA 8212

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| ||||| |||||||| || ||||| |||
Db       8213 CAGTCGCAAATGGCAATTCGCGTGACACAACGTTAA 8248


RESULT 9
ECCSGABDG
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
```

```
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           73.0%;  Score 332.8;  DB 1;  Length 4680;
  Best Local Similarity  83.1%;  Pred. No. 7.5e-87;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||
Db       3233 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 3292

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db       3293 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 3352

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db       3353 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 3412

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db       3413 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 3472

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db       3473 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 3532

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db       3533 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 3592

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       3593 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 3652

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db       3653 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 3688


RESULT 10
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
```

```
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRPPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

  Query Match           73.0%;  Score 332.8;  DB 1;  Length 10346;
  Best Local Similarity  83.1%;  Pred. No. 7.7e-87;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db       8497 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 8556

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || ||||||||||||| || |||
Db       8557 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 8616

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db       8617 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 8676

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db       8677 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 8736

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | ||||||  ||||  |||||||| |||| |||||||| ||  ||||||  |||||||
Db       8737 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 8796

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||  ||||||  | ||||| |||||||  |||    ||||||||||||||||
Db       8797 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 8856

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       8857 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 8916

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| ||||||||||||||| || ||||| |||
Db       8917 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 8952


RESULT 11
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                    Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .15047
                      /organism="Escherichia coli K12"
                      /mol_type="genomic DNA"
                      /strain="K12"
                      /db_xref="taxon:83333"
                      /map="23.7-24.0 min"
                      /clone="Kohara clone #231"
                      /note="Nucleotide position 1097101-1112147 from the
                      initiation site of ThrA (0 min.).~This clone is from
                      Kohara lambda miniset library."
     CDS              <1. .276
                      /note="unnamed protein product; ORF_ID:o230#5"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="BAA35822.1"
                      /db_xref="GI:4062605"
                      /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
                      WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     CDS              1152. .2090
                      /note="ORF_ID:o231#1
                      similar to SwissProt Accession Number P45637"
                      /codon_start=1
```

```
                /transl_table=11
                /product="Hypothetical 33.0 kd protein in proB-proA
                intergenic region."
                /protein_id="BAA35823.1"
                /db_xref="GI:4062606"
                /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
                WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
                AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
                PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
                GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
                AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
CDS             2145. .2882
                /note="unnamed protein product; ORF_ID:o231#2"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35824.1"
                /db_xref="GI:4062607"
                /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
                ILNVSPRRLLNFLESRGMAPIAEFADL"
CDS             2906. .3460
                /note="ORF_ID:o231#3
                similar to PIR Accession Number F64035"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1543"
                /protein_id="BAA35825.1"
                /db_xref="GI:4062608"
                /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                RTMAPLTRDAISAMWDELEEDSEE"
CDS             3514. .4053
                /note="unnamed protein product; ORF_ID:o231#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35826.1"
                /db_xref="GI:4062609"
                /translation="MCSISFMAAPSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                LAARSAKTAHREQEIKNKA"
gene            complement(4117. .4950)
                /gene="csgG"
CDS             complement(4117. .4950)
                /gene="csgG"
                /note="ORF_ID:o231#5
                similar to SwissProt Accession Number P52103"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgG precursor."
                /protein_id="BAA35827.1"
                /db_xref="GI:4062610"
                /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                S"
gene            complement(4977. .5393)
                /gene="csgF"
CDS             complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
gene            complement(5418. .5807)
                /gene="csgE"
CDS             complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
gene            complement(5812. .6462)
                /gene="csgD"
CDS             complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene            7217. .7672
                /gene="csgB"
CDS             7217. .7672
```

Query Match 73.0%; Score 332.8; DB 1; Length 15047;
Best Local Similarity 83.1%; Pred. No. 7.8e-87;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db       7217 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 7276

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db       7277 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 7336

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||||| ||||| ||||||||||  || ||  |||||||||     ||
Db       7337 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 7396

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db       7397 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 7456

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db       7457 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 7516

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| |||||||| |||     ||||||||||||||||||
Db       7517 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 7576

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       7577 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 7636

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
```

```
Db       7637 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 7672


RESULT 12
AE016981
LOCUS       AE016981               292504 bp    DNA     linear   BCT 22-APR-2003
DEFINITION  Shigella flexneri 2a str. 2457T section 4 of 16 of the complete
            genome.
ACCESSION   AE016981 AE014073
VERSION     AE016981.1  GI:30040616
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 2457T
  ORGANISM  Shigella flexneri 2a str. 2457T
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Complete Genome Sequence and Comparative Genomics of Shigella
            flexneri Serotype 2a Strain 2457T
  JOURNAL   Infect. Immun. 71 (5), 2775-2786 (2003)
   PUBMED   12704152
REFERENCE   2  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .292504
                     /organism="Shigella flexneri 2a str. 2457T"
                     /mol_type="genomic DNA"
                     /strain="2457T"
                     /serotype="2a"
                     /db_xref="taxon:198215"
     gene            complement(223. .441)
                     /gene="infA"
                     /locus_tag="S0884"
     CDS             complement(223. .441)
                     /gene="infA"
                     /locus_tag="S0884"
                     /function="factor; Proteins - translation and
                     modification"
                     /note="residues 1 to 72 of 72 are 100.00 pct identical to
                     residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
                     /codon_start=1
                     /transl_table=11
                     /product="protein chain initiation factor IF-1"
                     /protein_id="AAP16348.1"
                     /db_xref="GI:30040617"
                     /translation="MAKEDNIEMQGTVLETLPNTMFRVELENGHVVTAHISGKMRKNY
                     IRILTGDKVTVELTPYDLSKGRIVFRSR"
     gene            complement(726. .1430)
                     /gene="aat"
                     /locus_tag="S0885"
     CDS             complement(726. .1430)
                     /gene="aat"
                     /locus_tag="S0885"
                     /function="enzyme; Aminoacyl tRNA synthetases, tRNA
                     modification"
                     /note="residues 1 to 234 of 234 are 100.00 pct identical
                     to residues 1 to 234 of 234 from Escherichia coli K-12 :
                     B0885"
                     /codon_start=1
                     /transl_table=11
                     /product="leucyl, phenylalanyl-tRNA-protein transferase"
                     /protein_id="AAP16349.1"
                     /db_xref="GI:30040618"
                     /translation="MRLVQLSRHSIAFPSPEGALREPNGLLALGGDLSPARLLMAYQR
                     GIFPWFSPGDPILWWSPDPRAVLWPESLHISRSMKRFHKRSPYRVTMNYAFGQVIEGC
                     ASDREEGTWITRGVVEAYHRLHELGHAHSIEVWREDELVGGMYGVAQGTLFCGESMFS
                     RMENASKTALLVFCEEFIGHGGKLIDCQVLNDHTASLGACEIPRRDYLNYLNQMRLGR
                     LPNNFWVPRCLFSPQE"
     gene            complement(1472. .3193)
                     /gene="cydC"
                     /locus_tag="S0886"
     CDS             complement(1472. .3193)
                     /gene="cydC"
                     /locus_tag="S0886"
                     /function="transport; Transport of small molecules: Other"
                     /note="residues 1 to 573 of 573 are 97.03 pct identical to
                     residues 1 to 573 of 573 from Escherichia coli K-12 :
                     B0886"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-binding component of cytochrome-related
                     transport"
                     /protein_id="AAP16350.1"
                     /db_xref="GI:30040619"
                     /translation="MRALLPYLALYKRHKWMLSLGIVLAIVTLLASIGLLTLSGWFLS
                     ASAVAGVAGLYSFNYMLPAAGVRGAAITRTAGRYFERLVSHDATFRVLQHLRIYTFSK
                     LLPLSPAGLARYRQGELLNRVVADVDTLDHLYLRVISPLVGAFVVIMVVTIGLSFLDF
                     TLAFTLGSVMLLTLFLMPPLFYRAGKSTGQNLTHLRGQYRQQLTAWLQGQAELTIFGA
                     SDRYRTQLENTEIQWLEAQRRQSELTALSQAIMLLIGALAVILMLWMASGGVGGNAQP
                     GALIALFVFCALAAFEALAPVTGAFQHLGQVIASAVRISDLTDQKPEVTFPDTQTRVA
                     DRVSLTLRDVQFTYPEQSQQALKGISLQVNAGEHIAILGRTGCGKSTLLQLLTRAWDP
                     QQGEILLNDSPIASLNEAALRQTISVVPQRVHLFSATLRDNLLLASPGSSDEALSEIL
                     RRVGLEKLLEDAGLNSWLGEGGRQLSGGELRRLAIARTLLHDAPLVLLDEPTEGLDAT
                     TESQILELLAKMMREKTVLMVTHRLRGLSRFQQIIVMDNGQIIEQGTHAELLARQGRY
                     YQFKQGL"
     gene            complement(3194. .4960)
                     /gene="cydD"
                     /locus_tag="S0887"
     CDS             complement(3194. .4960)
                     /gene="cydD"
                     /locus_tag="S0887"
                     /function="transport; Not classified"
                     /note="residues 1 to 588 of 588 are 92.85 pct identical to
                     residues 1 to 588 of 588 from Escherichia coli K-12 :
                     B0887"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-binding component of cytochrome-related
                     transport"
                     /protein_id="AAP16351.1"
                     /db_xref="GI:30040620"
                     /translation="MNKSRQKELTRWLKQQSVISQRWLNISRLLGFVSGILIIAQAWF
                     MARILQHMIMENIPREALLLPFTLLVLTFVLRAWVVWLRERVGYHAGQHIRFAIRRQV
                     LDRLQQAGPAWIQGKPAGSWATLVLEQIDDMHDYYARYLPQMALAVSVPLLIVVAIFP
                     SNWAAALILLGTAPLIPLFMALVGMGAADANRRNFLALARLSGHFLDRLRGMETLRIF
                     GRGEAEIESIRSASEDFRQRTMEVLRLAFLSSGILEFFTSLSIALVAVYFGFSYLGEL
                     DFGHYDTGVTLAAGFLALILAPEFFQPLRDLGTFYHAKAQAVGAADSLKTFMETPLAH
                     PQRGEAELALTDPLTIEAEDLFITSPEGKTLAGPLNFTLPAGQRAVLVGRSGSGKSSL
                     LNALSGFLSYQGSLRINGIELRDLSPESWRKHLSWVGQNPQLPAAKLRDNVLLARPDA
                     SEQELQAALDNAWVSEFLPLLPQGVDTPVGDQAARLSVGQAQRVAVARALLNPCSLLL
                     LDEPAASLDAHSEQRVMEALNAASLRQTTLMVTHQLEDLADWDVIWVMQDGRIIEQGR
                     YAELSVAGGPFATLLAHRQEEI"
     gene            complement(5083. .6048)
                     /gene="trxB"
                     /locus_tag="S0888"
     CDS             complement(5083. .6048)
                     /gene="trxB"
                     /locus_tag="S0888"
                     /function="enzyme; 2-Deoxyribonucleotide metabolism"
                     /note="residues 1 to 321 of 321 are 95.95 pct identical to
                     residues 1 to 321 of 321 from Escherichia coli K-12 :
                     B0888"
                     /codon_start=1
                     /transl_table=11
                     /product="thioredoxin reductase"
                     /protein_id="AAP16352.1"
                     /db_xref="GI:30040621"
```

```
                     /translation="MGTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQL
                     TTTTEVENWPGDPNDLTGPLLMARMHEHATKFETEIIFDHINKVDLQNRPFRLNGDNG
                     EYTCDALIIATGASARYLGLPSEEAFKGRGVSACATCDGFFYRNQKVAVIGGGNTAVE
                     EALYLSNIASEVHLIHRRDGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGV
                     TGVRLRDTQNSDNIESLDVAGLFVAIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQ
                     TSIPGVFAAGDVMDHIYRQAITSAGTGCMAALDAERYLDGLADAK"
     gene            6592. .7086
                     /gene="lrp"
                     /locus_tag="S0889"
     CDS             6592. .7086
                     /gene="lrp"
                     /locus_tag="S0889"
                     /function="regulator; Central intermediary metabolism:
                     Pool, multipurpose conversions"
                     /note="residues 1 to 164 of 164 are 100.00 pct identical
                     to residues 1 to 164 of 164 from Escherichia coli K-12 :
                     B0889"
                     /codon_start=1
                     /transl_table=11
                     /product="regulator for leucine (or lrp) regulon and
                     high-affinity branched-chain amino acid transport system"
                     /protein_id="AAP16353.1"
                     /db_xref="GI:30040622"
                     /translation="MVDSKKRPGKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTP
                     CLERVRRLERQGFIQGYTALLNPHYLDASLLVFVEITLNRGAPDVFEQFNTAVQKLEE
                     IQECHLVSGDFDYLLKTRVPDMSAYRKLLGETLLRLPGVNDTRTYVVMEEVKQSNRLV
                     IKTR"
     gene            7221. .11249
                     /gene="ftsK"
                     /locus_tag="S0890"
     CDS             7221. .11249
                     /gene="ftsK"
                     /locus_tag="S0890"
                     /function="phenotype; Cell division"
                     /note="residues 1 to 737 of 1342 are 68.24 pct identical
                     to residues 1 to 737 of 1329 from Escherichia coli K-12 :
                     B0890"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein"
                     /protein_id="AAP16354.1"
                     /db_xref="GI:30040623"
                     /translation="MSQEYTEDKEVTLTKLSSGRRLLEALLILIVLFAVWLMAALLSF
                     NPSDPSWSQTAWHEPIHNLGGMPGAWLADTLFFIFGVMAYTIPVIIVGGCWFAWRHQS
                     SDEYIDYFAVSLRIIGVLALILTSCGLAAINADDIWYFASGGVIGSLLSTTLQPLLHS
                     SGGTIALLCVWAAGLTLFTGWSWVTIAEKLGGWILNILTFASNRTRRDDTWVDEDEYE
                     DDEEYEDENHGKQHESRRARILRGALARRKRLAEKFINPMGRQTDAALFSGKRMDDDE
                     EITYTARGVAADPDDVLFSGNRATQPEYDEYDPLLNSAPITEPVAVAAAATTATQSWA
                     APVEPVTQTPPVASVDVPPSQPTVAWQPVPGPQTGEPVIAPAPEGYPQQSQYAQPAVQ
                     YNEPLQQPVQPQQPYYAPAAEQPAQQPYYAPAPEQPVAGNAWQAEEQQSTFAPLSTYQ
                     TEQTYQQPAAQEPLYQQPQPVEQQPVVEPEPVVEETKPARPPLYYFEEVEEKRARERE
                     QLAAWYQPIPEPVKEPEPIKSSLKAPSVAAVPPVEAAAAVSPLASGVKKATLATGAAA
                     TVAAPVFSLANSGGPRPQVKEGIGPQLPRPKRIRVPTRRELASYGIKLPSQRAAEEKA
                     REAQRNQYDSGDQYNDDEIDAMQQDELARQFAQTQQQRYGEQYQHDVPVNAEDADAAA
                     EAELARQFAQTQQQRYSGEQPAGANPFSLDDFEFSPMKALLDDGPHEPLFTPIVEPVQ
                     QPQQPVAPQQQYQQPQQPVAPQQQYQQPQQQVAPQPQYQQPQQPVAPQQQYQQPQQPV
                     APQPQYQQPQQPVAPQPQYQQPQQPVAPQPQDTLLHPLLMRNGDSRPLHKPTTPLPSL
                     DLLTPPPSEVEPVDTFALEQMARLVEARLADFRIKADVVNYSPGPVITRFELNLAPGV
                     KAARISNLSRDLARSLSTVAVRVVEVIPGKPYVGLELPNKKRQTVYLREVLDNAKFRD

  Query Match           73.0%;  Score 332.8;  DB 1;  Length 292504;
  Best Local Similarity  83.1%;  Pred. No. 8.6e-87;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db     210815 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 210874

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db     210875 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 210934

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| |||||||| |  || ||  |||||||||     ||
Db     210935 TCTTCATTTAATCAGGCAGCCATAATTGGTCAGGCTGGGACTAATAATAGTGCTCAGTTA 210994

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | |||||||  | |  ||||||||
Db     210995 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAATCGGGCA 211054

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db     211055 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 211114

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db     211115 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 211174

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| |||||||||||||||| |
Db     211175 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAGTTGTAGTGCAGAGA 211234

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| ||||| |||||||| || ||||| |||
Db     211235 CAGTCGCAAATGGCAATTCGCGTGACACAACGTTAA 211270
```

```
RESULT 13
AE016759
LOCUS       AE016759               306358 bp    DNA     linear   BCT 24-DEC-2002
DEFINITION  Escherichia coli CFT073 section 5 of 18 of the complete genome.
ACCESSION   AE016759 AE014075
VERSION     AE016759.1  GI:26107527
KEYWORDS    .
SOURCE      Escherichia coli CFT073
  ORGANISM  Escherichia coli CFT073
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Extensive Mosiac Structure Revealed by the Complete Genome Sequence
            of Uropathogenic Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
   PUBMED   12471157
REFERENCE   2  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .306358
                     /organism="Escherichia coli CFT073"
                     /mol_type="genomic DNA"
                     /strain="CFT073"
                     /db_xref="taxon:199310"
     gene            complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
     CDS             complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
                     /note="Residues 27 to 1231 of 1245 are 80.16 pct identical
                     to residues  1 to 1209 of 1218 from GenPept.129 :
                     >emb|CAD05883.1| (AL627276) putative ABC transporter
                     protein [Salmonella enterica subsp. enterica serovar
                     Typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP binding cassette (ABC) transporter homolog"
                     /protein_id="AAN79727.1"
```

```
                /db_xref="GI:26107528"
                /translation="MRDRSVLAFLRNDPASGSYFTDGGGLMPANHTPTPAQSWIVRLA
                RVCWERKKLSVIVVVASVSTILLAALTPLLTRQAVNDALAGNPARLPWLACGLLLIAF
                FDFIGNYVRRGYAGMLSLWVQHTLRGRVFDSIQKLDGAGQDALRTGQVISRTNSDLQQ
                VHTLLQMCPVPLAVFTYYIAGIAVMLWMSPAMTLIVVCVLVCLAITALRARRRVFAQT
                GMASDQLANLTEHIREVLAQISVVKSCVAEMRETHWLDRQSRQIVRVRIGAVISQAMP
                GATMLALPVLGQIVLLCYGGWSVMHGRIDLGTFVAFASFLAMLTGPTRVLASFLVIAQ
                RTQASVERVFALIDTRSQMEDGTESINSQVVGLELENMSFDYHHGDRHILSDISFSLR
                AGETVAVVGASGSGKSTLLMLLARFYDPCSGKIWLNTSEGRQNLRDIRLEALRRRVGI
                VFEDAFLFAGTVAENIAYGHPQATADDIRRAAAAAGASDFINALPKGFDSLLTERGTN
                LSGGQRQRIALARALITAPDVLILDDTTSAVDAVTEAEINTALGRYADEGHMLLVIAR
                RRSTLQLASRVVVLDKGRMVDTGTPAELEARCPAFRALMTGDSDFLATSHNSHNELWP
                AEPATQDDVTDTGDKGFVARMTRVPENAVQQALAGKGRKVTSLLKPVAWMFVIAALLI
                ALDSAAGVGVLILLQHGIDSGVAAGDMSTIGLCALLALCLVIVGWCSYSLQTVFAARA
                AESVQHSVRLRSFGHMLRLGLPWHEKHADSRLTRMTVDVDSLARFLQNGLAGAATSLV
                TMFAIAATMFWLDPILALTALSAVPVAALATMIYRRLSTPAYAQARLEIGKVNSTLQE
                KVSGLRVVQSHGQQELEGARLRALSERFRATRVRAQKYLAVYFPFLTFCTEASYAAVL
                LVGASQVAAGEMTAGVLAAFFLLLGQFYGPVQQLSGIVDAWQQATASGKHIDELLATE
                GTENLGSSSVLPVTGALHLDEVTFSYPDSHEPALNKLTLTIPEGMVVAVVGRSGAGKS
                TLIKLIAGLYFPTHGNIRIGVQMLDDASLTEYRRQIGLVDQDVALFSSDIAENIRYSR
                PSATNEDVEIASQRAGLYEMVCNLPQGFRTPVNNGGADLSAGQRQLIALARAQLANAH
                ILLLDEATSCLDRTSEERLMSSLTDVVHAGKHSALIVAHRLTTAQRCDLIAVIDKGLL
                AEYGTHEQLLSAGGLYTRLWHDSVSSTALHRQHNMKEETPG"
gene            complement(3882. .5045)
                /gene="iroB"
                /locus_tag="c1254"
CDS             complement(3882. .5045)
                /gene="iroB"
                /locus_tag="c1254"
                /note="Escherichia coli O157:H7 ortholog: z1190"
                /codon_start=1
                /transl_table=11
                /product="Putative glucosyltransferase"
                /protein_id="AAN79728.1"
                /db_xref="GI:26107529"
                /translation="MRRLPDLDRQAERDFLMRILFVGPPLYGLLYPVLSLAQAFRVNG
                HEVLIASGGQFAQKAAEAGLVVFDAAPGLDSEAGYRHHEAQRKKSNIGTQMGNFSFFS
                EEMADHLVEFAGHWRPDLIIYPPLGVIGPLIAAKYDIPVVMQTVGFGHTPWHIRGVTR
                SLTDAYRRHNVGATPRDMAWIDVTPPSMSILENDGEPIIPMQYVPYNGGAVWEPWWER
                RPDRKRLLVSLGTVKPMVDGLDLIAWVMDSASEVDAEIILHISANARSDLRSLPSNVR
                LVDWIPMGVFLNGADGFIHHGGAGNTLTALHAGIPQIVFGQGADRPVNARVVAERGCG
                IIPGDVGLSSNMINAFLNNRSLRKASEEVAAEMAAQPCPGEVAKSLITMVQKG"
gene            5535. .5834
                /locus_tag="c1255"
CDS             5535. .5834
                /locus_tag="c1255"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79729.1"
                /db_xref="GI:26107530"
                /translation="MKPTLLQTVENICLQILFVLFLFTGTLIDVEALVSCSSVAKNVQ
                RTPASSVSSSVYSASVSRTAADIMLLSNNPLFAPGQVSFFGCSLAASGIHVTILE"
repeat_region   5867. .6577
                /insertion_seq="IS1541A-like"
gene            5961. .6470
                /locus_tag="c1256"
CDS             5961. .6470
                /locus_tag="c1256"
                /note="Residues 18 to 169 of 169 are 94.07 pct identical
                to residues  1 to 152 of 152 from SwissProt.40 :
                >sp|Q57334|T200_SALTY Transposase for insertion sequence
                element IS200"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79730.1"
                /db_xref="GI:26107531"
                /translation="MRSGNCKCQTRNQKGVPMGNEKSLAHTRWNCKYHIVFAPKYRRQ
                VFYREKRRAIGCILRKLCEWKSVRILEAECCADHIHMLVEIPPKMSVSGFMGYLKGKS
                SLMPYEQFGDLKPKYRNREFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPY
                PGSPPTGRK"
repeat_region   complement(6578. .6940)
                /note="incomplete"
                /insertion_seq="IS4"
gene            complement(6598. .6960)
                /locus_tag="c1257"
CDS             complement(6598. .6960)
                /locus_tag="c1257"
                /function="IS, phage, Tn"
                /note="Residues 8 to 120 of 120 are 98.23 pct identical to
                residues 316 to 428 of 442 from MG1655 : b4278"
                /codon_start=1
                /transl_table=11
                /product="Putative conserved protein"
                /protein_id="AAN79731.1"
                /db_xref="GI:26107532"
                /translation="MPRQPDQGYREIKQTMQLSRLTLRSKKPELVEQELWGVLLAYNL
                VRYQMIKMAEHLKGYWPNQLSFSESCGMVMRMLMTLQGASPGRIPELMRDLASMGQLV
                KLPTRRETAFPRVVKERP"
gene            7651. .8013
                /locus_tag="c1258"
CDS             7651. .8013
                /locus_tag="c1258"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79732.1"
                /db_xref="GI:26107533"
                /translation="MLGACFYSWGTSFYYGFPWWGAELVLMMWPEACFMLLQSSLYSL
                LDGEQVLFSFWELNKKQYEGFEFYQIFSGNIVASCSSCSGVSSNSSAHYARCTDFMRH
                YSLNNYAACQLGRQACLS"
gene            7835. .8269
                /locus_tag="c1259"
CDS             7835. .8269
                /locus_tag="c1259"
                /note="Residues 1 to 139 of 144 are 50.71 pct identical to
                residues  104 to 242 of 258 from GenPept.129 :
                >gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
                [Shigella flexneri]"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79733.1"
                /db_xref="GI:26107534"
                /translation="MKDLSFIRFFLAILLLLVPPALEFPVIHRHITPGALTLCVITAL
                IITLLASLGGRLVSVKCISEISFIRRHRIECMMAGFMIYFWTFSLIAGWYKPQFKKEY
                QTLHYENARYYVLARYDGRLVLSQSYRSGSRKFVIINGGHRL"
gene            8477. .8553
                /locus_tag="c5534"
tRNA            8477. .8553
                /locus_tag="c5534"
                /product="tRNA-Arg"
                /function="RNA; tRNA"
gene            complement(8832. .9290)
                /locus_tag="c1260"
CDS             complement(8832. .9290)
                /locus_tag="c1260"
                /function="putative enzyme; Transposases Insertion
                Sequence Associated"
                /note="Escherichia coli O157:H7 ortholog: z4324"
                /codon_start=1
                /transl_table=11
                /product="Putative Transposase"
                /protein_id="AAN79734.1"
                /db_xref="GI:26107535"
                /translation="MASAADARRIVSHYERRWLIEEYHKAWKSGGTCVESLRMQTRDN
                LERMVVIKAFIAVRMLGLRQEGISEETQNDSCKKILTPTEWKLLWVKLEGKQLPSQTP
                TLKWACLKLGRWHDSKRTGRPGWVVMWDGWFRLQDMVEGYPVMKSLDQEI"
gene            complement(9302. .9667)
                /locus_tag="c1261"
CDS             complement(9302. .9667)
                /locus_tag="c1261"
                /note="Residues 1 to 35 of 121 are 97.14 pct identical to
                residues  59 to 93 of 316 from GenPept.129 :
                >gb|AAK18437.1|AF348706_126 (AF348706) putative
```

```
                     transposase [Shigella flexneri]"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79735.1"
                     /db_xref="GI:26107536"
                     /translation="MAERLGEIQKRVITVCDREADIWHYLHYKVSHGQRFVVRTAQNS

  Query Match             73.0%;  Score 332.8;  DB 1;  Length 306358;
  Best Local Similarity   83.1%;  Pred. No. 8.6e-87;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db      44773 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 44832

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| |||||||||||||||| || |||
Db      44833 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTTGCGGTAAATGAATTGAGTAAG 44892

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db      44893 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 44952

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db      44953 CGGCAGGGAGGCTCAAAACTTTTGACGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 45012

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | || ||| ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db      45013 AAGATTGATCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 45072

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db      45073 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 45132

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db      45133 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 45192

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db      45193 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 45228


RESULT 14
AF275733
LOCUS       AF275733               1711 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
            CsgA protein (csgA) genes, complete cds.
ACCESSION   AF275733
VERSION     AF275733.1  GI:14039399
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Mutations in the csgD promoter associated with variations in curli
            expression in certain strains of Escherichia coli O157:H7
  JOURNAL   Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
  MEDLINE   21218556
   PUBMED   11319125
REFERENCE   2  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
            Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES             Location/Qualifiers
     source          1. .1711
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="43895 Red Variant"
                     /serotype="O157:H7"
                     /db_xref="taxon:562"
     gene            757. .1212
                     /gene="csgB"
     CDS             757. .1212
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="AAK53211.1"
                     /db_xref="GI:14039400"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1253. .1711
                     /gene="csgA"
     CDS             1253. .1711
                     /gene="csgA"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="AAK53212.1"
                     /db_xref="GI:14039401"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match             72.6%;  Score 331.2;  DB 1;  Length 1711;
  Best Local Similarity   82.9%;  Pred. No. 2.2e-86;
  Matches  378;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db        757 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 816

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| |||||||||||||||| || |||
Db        817 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 876

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db        877 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 936

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db        937 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 996

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db        997 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 1056

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db       1057 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 1116

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db       1117 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 1176

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db       1177 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 1212


RESULT 15
AE005315
LOCUS       AE005315               10190 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
            of 155.
ACCESSION   AE005315 AE005174
VERSION     AE005315.1  GI:12514572
```

```
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1..10190
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            617..1072
                     /gene="csgB"
                     /note="synonym: Z1675"
     CDS             617..1072
                     /gene="csgB"
                     /function="structural component; Cell exterior
                     constituents: Surface structures"
                     /note="Residues 1 to 151 of 151 are 100.00 pct identical
                     to residues 1 to 151 of 151 from Escherichia coli K-12
                     Strain MG1655: B1041"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAG55787.1"
                     /db_xref="GI:12514573"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1113..1571
                     /gene="csgA"
                     /note="synonym: Z1676"
     CDS             1113..1571
                     /gene="csgA"
                     /function="regulator; Cell exterior constituents: Surface
                     structures"
                     /note="Residues 1 to 152 of 152 are 96.71 pct identical to
                     residues 1 to 151 of 151 from Escherichia coli K-12 Strain
                     MG1655: B1042"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin major subunit, coiled surface structures;
                     cryptic"
                     /protein_id="AAG55788.1"
                     /db_xref="GI:12514574"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
     gene            1630..1962
                     /gene="csgC"
                     /note="synonym: Z1677"
     CDS             1630..1962
                     /gene="csgC"
                     /function="putative structure; Not classified"
                     /note="Residues 1 to 110 of 110 are 98.18 pct identical to
                     residues 1 to 110 of 110 from Escherichia coli K-12 Strain
                     MG1655: B1043"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="AAG55789.1"
                     /db_xref="GI:12514575"
                     /translation="MNALLLLAALSSQITFNTTQQGDMYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     gene            2083..2394
                     /gene="Z1678"
     CDS             2083..2394
                     /gene="Z1678"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 103 of 103 are 98.05 pct identical to
                     residues 1 to 103 of 103 from Escherichia coli K-12 Strain
                     MG1655: B1044"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55790.1"
                     /db_xref="GI:12514576"
                     /translation="MFRPFLNSLMLGSLFFPFIAIAGSTAQGGVIHFYGQIVEPACDV
                     STQSSPVEMNCPQNGSVPGKTYSSKALMSGNVKNAQIASVKVQYLDKQKKLAVMNIEY
                     N"
     gene            2489..3022
                     /gene="Z1679"
     CDS             2489..3022
                     /gene="Z1679"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 177 of 177 are 100.00 pct identical
                     to residues 1 to 177 of 177 from Escherichia coli K-12
                     Strain MG1655: B1045"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyprotein"
                     /protein_id="AAG55791.1"
                     /db_xref="GI:12514577"
                     /translation="MKTRIHVVQGDITKLAVDVIVNAANPSLMGGGGVDGAIHRAAGP
                     ALLDACLKVRQQQGDCPTGHAVITLAGDLPAKAVVHTVGPVWRGGEQNEDQLLQDAYL
                     NSLRLVAANSYTSVAFPAISTGVYGYPRAAAAEIAVKTVSEFITRHALPEQVYFVCYD
                     EENAHLYERLLTQQGDE"
     gene            2964..4445
                     /gene="ymdC"
                     /note="synonym: Z1680"
     CDS             2964..4445
                     /gene="ymdC"
                     /function="putative enzyme; Not classified"
                     /note="Residues 1 to 493 of 493 are 99.59 pct identical to
                     residues 1 to 493 of 493 from Escherichia coli K-12 Strain
                     MG1655: B1046"
                     /codon_start=1
                     /transl_table=11
                     /product="putative synthase"
                     /protein_id="AAG55792.1"
                     /db_xref="GI:12514578"
                     /translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPGQCGLFPL
                     EKSLDAFAARYRLAEMSEHTLDVQYYIWQDDMSGRLLFSALLAAAKRGVRVRLLLDDN
                     NTPGLDDILRLLDSHPRIEVRLFNPFSFRLLRPLGYITDFSRLNRRMHNKSFTVDGVV
                     TLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWYCKSVSPLQQVLDV
                     PEGEMADRIELPASWHNDAMTHRYLRKMESSPFINHLVDGTLPLIWAKTRLLSDDPAK
                     GEGKAKRHSLLPQRLFDIMGSPSERIDIISSYFVPTRAGVAQLLRMVRKXVKIAILTN
                     SLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREQSSTLHDRGITGNSGASLHAKT
                     FSIDGKTVFIGSFNFDPRSTLLNTEMGFVIESETLAQLIDKRFIQSQYDAAWQLRLDR
                     WGRINWVDRHAKKEIVLKKEPATSFWKRVMVRLASILPVEWLL"
     gene            complement(4453..5610)
                     /gene="ymdD"
```

```
                     /note="synonym: Z1681"
     CDS             complement(4453. .5610)
                     /gene="ymdD"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 385 of 385 are 99.22 pct identical to
                     residues 1 to 385 of 385 from Escherichia coli K-12 Strain
                     MG1655: B1047"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55793.1"
                     /db_xref="GI:12514579"
                     /translation="MNPVPAQREYFXDSIRAWLMLLGIPFHISLIYSSHTWHVNSAEP
                     SLWLTLFNDFIHSFRMQVFPVISGYFSYMLFLRYPLKKWWKVRVERVGIPMLTAIPLL
                     TLPQFIMLQYVKGKAESWPGLSLYDKYNTLAWELISHLWFLLVLVVMTTLCVWIFKRI
                     RNNLENSDKTNKKFSMVKLSVIFLCLGIGYAVIRRTIFIVYPPILSNGTFNFIVMQTL
                     FYLPFFILGALAFIFPHLKALFTTPSRGCTLAAALAFVAYLLNQRYGSGDAWMYETES
                     VITMVLGLWMVNVVFSFGHRLLNFQSARVTYFVNASLFIYLVHHPLTLFFGAYITPHI
                     TSNWLGFLCGLIFVVGIAIILYEIHLRIPLLKFLFSGKPVVKRENDKAPAR"
     gene            6003. .7538
                     /gene="mdoG"
                     /note="synonym: Z1683"
     CDS             6003. .7538
                     /gene="mdoG"
                     /function="enzyme; Osmotic adaptation"
                     /note="Residues 1 to 511 of 511 are 99.80 pct identical to
                     residues 1 to 511 of 511 from Escherichia coli K-12 Strain
                     MG1655: B1048"
                     /codon_start=1
                     /transl_table=11
                     /product="periplasmic glucans biosynthesis protein"
                     /protein_id="AAG55794.1"
                     /db_xref="GI:12514580"
                     /translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAQSLAGKGYEAPK
                     SNLPSVFRDMKYADYQQIQFNHDKAYWNNLKTPFKLEFYHQGMYFDTPVKINEVTATA
                     VKRIKYSPDYFTFGDVQHDKDTVKDLGFAGFKVLYPINSKDKNDEIVSMLGASYFRVI
                     GAGQVYGLSARGLAIDTALPSGEEFPRFKEFWIERPKPTDKRLTIYALLDSPRATGAY
                     KFVVMPGRDTVVDVQSKIYLRDKVGKLGVAPLTSMFLFGPNQPSPANNYRPELHDSNG
                     LSIHAGNGEWIWRPLNNPKHLAVSSFSMENPQGFGLLQRGRDFSRFEDLDDRYDLRPS
                     AWVTPKGEWGKGSVELVEIPTNDETNDNIVAYWTPDQLPEPGKEMNFKYTITFSRDED
                     KLHAPDNAWVQQTRRSTGDVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDTPVTAQTS
                     IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDAKKTTEMRAALVNADQTLSETWSYQLP
                     ANE"
     gene            7531. .10074
                     /gene="mdoH"
                     /note="synonym: Z1684"
     CDS             7531. .10074
                     /gene="mdoH"
                     /function="enzyme; Osmotic adaptation"

  Query Match           72.6%;  Score 331.2;  DB 1;  Length 10190;
  Best Local Similarity  82.9%;  Pred. No. 2.3e-86;
  Matches  378;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db        617 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 676

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  ||||||| | |||  |||||||||||| || ||||||||||||| || |||
Db        677 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 736

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              |||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db        737 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 796

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db        797 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 856

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||  |||||| ||||||
Db        857 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 916

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||     ||||||||||||||||||
Db        917 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 976

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db        977 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 1036

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db       1037 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 1072


Search completed: March 18, 2004, 15:02:27
Job time : 2314.32 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64618	Aac64618 Salmonell
	2	332.8	73.0	456	3	AAC64620	Aac64620 Escherich
	3	74.4	16.3	100	7	ACD68810	Acd68810 E. coli K
	4	66.4	14.6	100	7	ACD68812	Acd68812 E. coli K
	5	64.8	14.2	100	7	ACD68811	Acd68811 E. coli K
	6	37.6	8.2	456	3	AAC64623	Aac64623 AgfA::PT3
c	7	37.6	8.2	2000	7	ADA71938	Ada71938 Rice gene
	8	35.8	7.9	39	3	AAC64608	Aac64608 AgfA (SEF
	9	35	7.7	738	6	AAD32400	Aad32400 Chlamydia
	10	35	7.7	105325	6	ABK94407	Abk94407 DNA encod
	11	34.4	7.5	2006	9	ADC10081	Adc10081 Human NOV
c	12	34.4	7.5	18540	8	ACD26496	Acd26496 Human tum
	13	34.2	7.5	456	3	AAC64629	Aac64629 AgfA::PT3
	14	34	7.5	361	2	AAQ73066	Aaq73066 AgfA sequ
	15	34	7.5	361	2	AAT74141	Aat74141 Salmonell
	16	34	7.5	456	2	AAQ87467	Aaq87467 AgfA sequ
	17	34	7.5	456	2	AAT74142	Aat74142 Salmonell
	18	34	7.5	456	3	AAC64628	Aac64628 AgfA::PT3
	19	34	7.5	456	3	AAC64622	Aac64622 AgfA::PT3
	20	34	7.5	456	3	AAC64617	Aac64617 Salmonell
	21	34	7.5	456	3	AAC64625	Aac64625 AgfA::PT3
	22	34	7.5	456	3	AAC64624	Aac64624 AgfA::PT3
	23	34	7.5	2000	7	ADA71938	Ada71938 Rice gene
c	24	33.8	7.4	203654	7	ABX16034	Abx16034 Human gen
c	25	33.4	7.3	53585	2	AAX20251	Aax20251 Borrelia
	26	33.2	7.3	535	9	ADD29528	Add29528 Mouse cae
	27	33	7.2	33	3	AAC64614	Aac64614 S. enteri
	28	33	7.2	3699	6	ABN70306	Abn70306 Streptoco
	29	33	7.2	3699	6	ABN70834	Abn70834 Streptoco
	30	33	7.2	3717	6	ABN66453	Abn66453 Streptoco
	31	33	7.2	110000	6	ABN71527_03	Continuation (4 of
	32	32.8	7.2	418	3	AAH30547	Aah30547 Human col
	33	32.8	7.2	1083	9	ADC10077	Adc10077 Human NOV
	34	32.8	7.2	1098	9	ADC10079	Adc10079 Human NOV
	35	32.8	7.2	1105	9	ADC10075	Adc10075 Human NOV
	36	32.8	7.2	1168	9	ADC10069	Adc10069 Human NOV
	37	32.8	7.2	1171	9	ADC10071	Adc10071 Human NOV
	38	32.8	7.2	1299	9	ADC10073	Adc10073 Human NOV
	39	32.8	7.2	1938	9	ADC10067	Adc10067 Human NOV
	40	32.8	7.2	2649	7	ADA53932	Ada53932 Human cod
	41	32.8	7.2	2843	7	ACC44405	Acc44405 Gene enco
	42	32.4	7.1	779	1	AAN40272	Aan40272 Sequence
	43	32.4	7.1	1914	4	AAS51585	Aas51585 Staphyloc
	44	32.4	7.1	1929	4	AAS54408	Aas54408 Staphyloc
	45	32.4	7.1	10732	3	AAA10594	Aaa10594 Gene enco

ALIGNMENTS

RESULT 1
AAC64618
ID AAC64618 standard; DNA; 456 BP.
XX
AC AAC64618;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis AgfB DNA sequence SEQ ID NO:2.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36342.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

```
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 149 A; 87 C; 109 G; 111 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-135;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456


RESULT 2
AAC64620
ID   AAC64620 standard; DNA; 456 BP.
XX
AC   AAC64620;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Escherichia coli CsgB DNA sequence SEQ ID NO:4.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Escherichia coli.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36344.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 146 A; 78 C; 113 G; 119 T; 0 U; 0 Other;

  Query Match             73.0%;  Score 332.8;  DB 3;  Length 456;
  Best Local Similarity   83.1%;  Pred. No. 1.5e-95;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  |||||||| | |||  |||||||||||| || |||||||||||||| || |||
Db         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||| |||| ||||||||||  || ||  ||||||||||     ||
Db        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
```

```
                ||||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| ||||||||||| |
Db        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456


RESULT 3
ACD68810
ID   ACD68810 standard; DNA; 100 BP.
XX
AC   ACD68810;
XX
DT   18-SEP-2003  (first entry)
XX
DE   E. coli K12 MG1655 biochip probe SEQ ID 80.
XX
KW   Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS   Escherichia coli.
XX
PN   EP1260592-A1.
XX
PD   27-NOV-2002.
XX
PF   17-MAY-2001; 2001EP-00112179.
XX
PR   17-MAY-2001; 2001EP-00112179.
XX
PA   (MWGB-) MWG-BIOTECH AG.
XX
PI   Donner H,  Drescher B,  Huber A,  Weber J;
XX
DR   WPI; 2003-241155/24.
XX
PT   Biochip containing probes complementary with open reading frames in
PT   Escherichia coli K12, useful for detecting gene expression and expression
PT   patterns.
XX
PS   Claim 3; Page 23; 2004pp; German.
XX
CC   This invention describes a novel biochip comprising probe spots, each
CC   containing many identical probes. The probes are nucleotide sequences of
CC   30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC   least one includes a segment of at least 20 bases identical with, or
CC   complementary to, a segment of an open reading frame (orf) of Escherichia
CC   coli K12. The biochip is used for specific detection of gene expression
CC   in K12 and for determining the gene expression pattern, e.g. for
CC   diagnostic determination of which E. coli strains are present in the gut,
CC   and to determine the effects of e.g. growth media on gene expression. The
CC   biochip provides as comprehensive as possible detection of the K12
CC   genome, with simultaneous analysis of many different genes with a single
CC   device, and comparison of gene expression between K12 and its mutants or
CC   other E. coli strains in a single experiment. Apart from qualitative and
CC   quantitative information about gene expression, it also allows
CC   measurements of population densities for the various strains. The use of
CC   synthetic oligonucleotides for preparation of probes allows free
CC   variation in probe length and ensures high purity (and thus selectivity,
CC   reactivity and reproducibility); also synthetic probes are generally
CC   shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC   ACD81540 represent oligonucleotide probes used with the biochip described
CC   in the invention
XX
SQ   Sequence 100 BP; 27 A; 19 C; 26 G; 28 T; 0 U; 0 Other;

  Query Match             16.3%;  Score 74.4;  DB 7;  Length 100;
  Best Local Similarity   84.0%;  Pred. No. 1.6e-13;
  Matches   84;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        266 ACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTT 325
              || |||| |||||||| ||  |||||| |||||||||||||||||| |||||| | ||||
Db          1 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 60

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAA 365
              | ||||||| |||    |||||||||||||||||||| ||
Db         61 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 100


RESULT 4
ACD68812
ID   ACD68812 standard; DNA; 100 BP.
XX
AC   ACD68812;
XX
DT   18-SEP-2003  (first entry)
XX
DE   E. coli K12 MG1655 biochip probe SEQ ID 82.
XX
KW   Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS   Escherichia coli.
XX
PN   EP1260592-A1.
XX
PD   27-NOV-2002.
XX
PF   17-MAY-2001; 2001EP-00112179.
XX
PR   17-MAY-2001; 2001EP-00112179.
XX
PA   (MWGB-) MWG-BIOTECH AG.
XX
PI   Donner H,  Drescher B,  Huber A,  Weber J;
XX
DR   WPI; 2003-241155/24.
XX
PT   Biochip containing probes complementary with open reading frames in
PT   Escherichia coli K12, useful for detecting gene expression and expression
PT   patterns.
XX
PS   Claim 3; Page 23; 2004pp; German.
XX
CC   This invention describes a novel biochip comprising probe spots, each
CC   containing many identical probes. The probes are nucleotide sequences of
CC   30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC   least one includes a segment of at least 20 bases identical with, or
CC   complementary to, a segment of an open reading frame (orf) of Escherichia
CC   coli K12. The biochip is used for specific detection of gene expression
CC   in K12 and for determining the gene expression pattern, e.g. for
CC   diagnostic determination of which E. coli strains are present in the gut,
CC   and to determine the effects of e.g. growth media on gene expression. The
CC   biochip provides as comprehensive as possible detection of the K12
CC   genome, with simultaneous analysis of many different genes with a single
CC   device, and comparison of gene expression between K12 and its mutants or
CC   other E. coli strains in a single experiment. Apart from qualitative and
CC   quantitative information about gene expression, it also allows
CC   measurements of population densities for the various strains. The use of
CC   synthetic oligonucleotides for preparation of probes allows free
CC   variation in probe length and ensures high purity (and thus selectivity,
CC   reactivity and reproducibility); also synthetic probes are generally
CC   shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC   ACD81540 represent oligonucleotide probes used with the biochip described
CC   in the invention
XX
SQ   Sequence 100 BP; 28 A; 18 C; 27 G; 27 T; 0 U; 0 Other;

  Query Match             14.6%;  Score 66.4;  DB 7;  Length 100;
  Best Local Similarity   79.0%;  Pred. No. 5.6e-11;
  Matches   79;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        123 TTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACG 182
```

```
Db          1 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 60

Qy        183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAA 222
              |||| ||| |||||||| ||| | ||| || | ||||||
Db         61 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 100
```

RESULT 5
ACD68811
ID ACD68811 standard; DNA; 100 BP.
XX
AC ACD68811;
XX
DT 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 81.
XX
KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EP1260592-A1.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MWGB-) MWG-BIOTECH AG.
XX
PI Donner H, Drescher B, Huber A, Weber J;
XX
DR WPI; 2003-241155/24.
XX
PT Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX
PS Claim 3; Page 23; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SQ Sequence 100 BP; 31 A; 20 C; 29 G; 20 T; 0 U; 0 Other;

Query Match 14.2%; Score 64.8; DB 7; Length 100;
Best Local Similarity 78.0%; Pred. No. 1.8e-10;
Matches 78; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy        208 GTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAAC 267
              ||| || | ||||||||  | |  || ||||| ||  | ||||||| ||||  |||||||||
Db          1 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 60

Qy        268 TTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCA 307
              |||| |||||||| ||  |||||| ||||||||||||||
Db         61 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 100
```

RESULT 6
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

```
  Query Match             8.2%;  Score 37.6;  DB 3;  Length 456;
  Best Local Similarity   51.2%;  Pred. No. 0.15;
  Matches   88;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
              ||   |||    || |  ||  |||  ||||  |||  | ||   ||  | ||  |    ||
Db        203 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 262

Qy        332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTA 391
              || |    |  | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACG 322

Qy        392 CGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGT 443
              |  | ||  | |   ||   ||      || |    ||  ||| ||  || ||
Db        323 CTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGT 374


RESULT 7
ADA71938/c
ID   ADA71938 standard; DNA; 2000 BP.
XX
AC   ADA71938;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Rice gene, SEQ ID 5263.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
OS   Oryza sativa.
XX
PN   WO2003000898-A1.
XX
PD   03-JAN-2003.
XX
PF   22-JUN-2001; 2001WO-IB001105.
XX
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H,  Chen W,  Cooper B,  Glazebrook J,  Goff SA,  Hou Y;
PI   Katagiri F,  Quan S,  Tao Y,  Whitham S,  Xie Z,  Zhu T,  Zou G;
XX
DR   WPI; 2003-175290/17.
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction. (M1) is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

  Query Match             8.2%;  Score 37.6;  DB 7;  Length 2000;
  Best Local Similarity   12.8%;  Pred. No. 0.26;
  Matches   57;  Conservative  195;  Mismatches  189;  Indels    3;  Gaps    2;

Qy         14 TGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACAAATTATG 73
              |||    |  |::::|: ::: : :   ::  :::|  : |::  : :: :|:|:   ::

Db       1027 TGTAGGTGCTKMWRTYTRSMSWTYAMWKKYTKYMTAYSSTWKWYWAYKWRAYAWSRSRKT 968

Qy         74 ATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATC 133
              ::||||  : |:|   ::|  :: :      :      :::: :|:: :::  :  :
Db        967 WWCTGGKRMATYCGTKMWAAGRWRWRMAWCWYCCMWKWWKMTSCMWWKYWRTWSCWYTMW 908

Qy        134 AGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGAT 193
              :| ::  :   :  : |: ::: : ::    :|:|: | :|:   ::     |  :: : |
Db        907 WGAMRYAYYAMRRRRWTYKWSWRRMYWTMTKWAWTWMTCMCMAKWYMATGWATWMWWRYT 848

Qy        194 CAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGG 253
              :   : :|  :|  : | |: :  :  | :| :::: |:   :  : ::   :  : :
Db        847 MYTYCYAMTCAKCKYKMAMTKWWTTWACAWRATSWRWRAMAGMRWKRYKMKRAYWWRWWR 788

Qy        254 CAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATAT 313
              |:: |:| ::: :: ::  ::||:::    |:    : ::::::::: :  | ::  |:::
Db        787 CWKAGWARWMKSRYRWKWKKYATRYYWKMWAMTWWWSWRRWKSYRMWSGMGRMRWSAWRY 728

Qy        314 CGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCA 373
              |:  :::|  |:  :: ::| : :    ::  ::  :: :  :: :::: ::|:|: : :
Db        727 CS--RMKCAKTKYASSARWTKRAKRSYRYRRRWYWKRKGWTYRYRYWRSCRMTRARMSKR 670

Qy        374 ATATTACCCAGTACGGTACGCA-GAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATG 432
              :::  |::   ::: : | ::: :    :|:| ::|:::  ::::  ::  :::::  :
Db        669 RKWAGASMKSCWMYWRGARSMWYSKYSCSAKCCKKTRYMTSSYMSTGMYGMYSSYKSMSW 610

Qy        433 GCTATTCGCGTCACCCAACGCTAA 456
               ::: ::|  || ::  : |:|::
Db        609 TSKMSYMGKMTCTMYTSMKGSTRR 586


RESULT 8
AAC64608
ID   AAC64608 standard; DNA; 39 BP.
XX
AC   AAC64608;
XX
DT   06-AUG-2003  (revised)
DT   26-FEB-2001  (first entry)
XX
DE   AgfA (SEF17) recombinant agfA::PT3a generating PCR primer 17-A.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; PCR primer; ss.
XX
OS   Salmonella enteritidis.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Example 2; Page 61; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
```

```
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC   correct OS field.)
XX
SQ   Sequence 39 BP; 13 A; 7 C; 10 G; 9 T; 0 U; 0 Other;

  Query Match             7.9%;  Score 35.8;  DB 3;  Length 39;
  Best Local Similarity   94.9%;  Pred. No. 0.21;
  Matches   37;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        393 GCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATAT 431
              ||||||  |||||||||||||||||||||||||||||||
Db          1 GCAGAATTCAGCAGTTGTAGTGCAGAAACAGTCGCATAT 39


RESULT 9
AAD32400
ID   AAD32400 standard; DNA; 738 BP.
XX
AC   AAD32400;
XX
DT   29-AUG-2003  (revised)
DT   18-JUN-2002  (first entry)
XX
DE   Chlamydia pneumoniae nlpD gene.
XX
KW   Chlamydiaceae family; chronic infection; persistent infection; pyk; nlpD;
KW   Cpn0585; regulatory pathway; biosynthetic pathway; ompA; ompB; hsp60;
KW   lipopolysaccharide; cardiovascular system; respiratory tract; therapy;
KW   genital tract; reproductive system; atherosclerotic tissue; macrophage;
KW   multiple sclerosis; conjunctiva; prophylaxis; antibacterial; gene; ds.
XX
OS   Chlamydophila pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .738
FT                   /*tag=  a
FT                   /product= "Chlamydia pneumoniae nlpD protein"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200214516-A1.
XX
PD   21-FEB-2002.
XX
PF   17-AUG-2001; 2001WO-AU001021.
XX
PR   18-AUG-2000; 2000AU-00009540.
XX
PA   (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
PA   (MATH/) MATHEWS S A.
XX
PI   Timms P;
XX
DR   WPI; 2002-269197/31.
DR   P-PSDB; AAE20289.
XX
PT   Detecting Chlamydial organism in its persistent phase by detecting
PT   expression change of range of genes belonging to their respective
PT   biosynthetic pathways when expression is compared to that of organism in
PT   lytic phase.
XX
PS   Disclosure; Page 117-118; 196pp; English.
XX
CC   The invention relates to composition and methods for detecting organisms
CC   of the Chlamydiaceae family, including species of Chlamydophila and
CC   Chlamydia, in the persistent phase of their developmental cycle and for
CC   the diagnosis of chronic or persistent infections caused by such
CC   organisms. The composition is useful for modulating the expression of
CC   gene such as pyk, nlpD, Cpn0585, a gene belonging to same regulatory/
CC   biosynthetic pathway and ompA, ompB, hsp60, a gene involved in
CC   lipopolysaccharide biosynthesis. It is also useful for modulating the the
CC   level and/or functional activity of an expression product of these genes,
CC   where the gene is present in an epithelial cell (selected from
CC   cardiovascular system, respiratory tract, genital tract, reproductive
CC   system or conjunctiva), macrophage, or a cell associated with
CC   atherosclerotic tissue or associated with multiple sclerosis brain
CC   tissue. The composition is useful for treatment and/or prophylaxis of a
CC   chronic infection caused by an organism of the Chlamydiaceae family in a
CC   patient. Antigen associated with the persistent phase of the
CC   developmental cycle of an organism of the Chlamydiaceae family, is useful
CC   in the manufacture of a medicament, for treating and/or preventing
CC   Chlamydiaceae infection in a patient. The present sequence is Chlamydia
CC   pneumoniae nlpD gene. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ   Sequence 738 BP; 231 A; 158 C; 166 G; 183 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 35;  DB 6;  Length 738;
  Best Local Similarity   57.9%;  Pred. No. 1.2;
  Matches   62;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        340 GCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAA 399
              |||  ||   |||| |  | | || | | |||  ||||| |  |     |  | |   | |
Db          1 GCTGTTGATCAGAGAAATGCAGGCAGTGAGGTAAATATGAATCGTAGAGACATGGTAATA 60

Qy        400 ACAGCAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCAC 446
              ||||| || |||||| |       | || | ||||| ||  ||||||
Db         61 ACAGCTGTCGTAGTGAATGCTATATTGCTTGTGGCTCTTTTCGTCAC 107


RESULT 10
ABK94407
ID   ABK94407 standard; DNA; 105325 BP.
XX
AC   ABK94407;
XX
DT   27-AUG-2002  (first entry)
XX
DE   DNA encoding endothelin-3 (EDN-3).
XX
KW   Endothelin 3; EDN-3; endothelin converting enzyme; ECE;
KW   endothelin receptor; EDNR; signaling system; cardiovascular disease;
KW   coronary heart disease; hypertension; atherosclerosis; angiogenesis;
KW   fatty acid metabolism; diabetes; familial hypercholesterolaemia;
KW   forensic marker; transgenic animal; solid support; SNP;
KW   cardiovascular regulator; gene; ds; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(40215,A)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(59430,C)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63843,T)
```

```
FT                  /*tag= c
FT                  /standard_name= "Single nucleotide polymorphism"
XX
PN   WO200224747-A2.
XX
PD   28-MAR-2002.
XX
PF   31-AUG-2001; 2001WO-EP010087.
XX
PR   19-SEP-2000; 2000EP-00120123.
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI   Brinkmann U,  Hoffmeyer S;
XX
DR   WPI; 2002-435060/46.
XX
PT   Novel polynucleotide of the endothelin/endothelin converting
PT   enzyme/receptors of endothelin and endothelin converting enzyme signaling
PT   system associated with cardiovascular disease, useful for treating the
PT   disease.
XX
PS   Claim 1; Page; 190pp; English.
XX
CC   The invention describes a polynucleotide (I) of the endothelin
CC   (EDN)/endothelin converting enzyme (ECE)/receptors of EDN and ECE (EDNR)
CC   signaling system which is associated with a cardiovascular disease. (I),
CC   the gene encoding EDN, ECE or EDNR (II) or a vector (III) expressing (I)
CC   or (II) is useful for producing cells capable of expressing a molecular
CC   variant polypeptide which is associated with a cardiovascular disease.
CC   (II), (III), the EDN, ECE or EDNR polypeptide, or a cell expressing a
CC   molecular variant gene comprising (I) is useful for identifying and
CC   obtaining a pro-drug or drug capable of modulating the activity of a
CC   molecular variant of a polypeptide of the EDN/EDNR/ECE signaling system
CC   or its gene product, or for identifying and obtaining an inhibitor of the
CC   activity of a molecular variant of a polypeptide of the EDN/EDNR/ECE
CC   signaling system or its gene product. The isolated proteins and
CC   polynucleotides encoding them are useful for preparation of a
CC   pharmaceutical composition for treating a cardiovascular disease such as
CC   coronary heart disease, hypertension, atherosclerosis, or related to
CC   abnormal angiogenesis or fatty acid metabolism e.g. diabetes and familial
CC   hypercholesterolaemia. The gene or a polynucleotide fragment of the
CC   EDN/ECE/EDNR signaling system are useful as forensic markers, for
CC   creating a transgenic animal and in creation of a solid support
CC   comprising polynucleotides, genes, vectors, polypeptides, antibodies or
CC   host cells of the invention. This sequence encodes the cardioavscular
CC   regulator Endothelin-3 (EDN-3). Note: This sequence does not appear in
CC   the specification but has been obtained from GenBank using information
CC   given in the invention
XX
SQ   Sequence 105325 BP; 26263 A; 24618 C; 25835 G; 28609 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 35;  DB 6;  Length 105325;
  Best Local Similarity   56.5%;  Pred. No. 8.2;
  Matches   65;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||  || |       || ||  ||||   |  || || |||  |  ||  |||
Db      66519 ATGAAAAAGCAAATTAAGAGCATAATCCTGACTGCATGGGATGTGCCCAGCCTTCTAACA 66578

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAA 115
              |  ||||| | |   |||||   |||| | |||   ||        ||| ||||||
Db      66579 GGAACAAACTCTTTCCTGGCCTTTTCAAAGTATCTGTTACTATGTAATAAATTAA 66633


RESULT 11
ADC10081
ID   ADC10081 standard; DNA; 2006 BP.
XX
AC   ADC10081;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human NOVX polypeptide coding sequence SEQ ID NO: 101.
XX
KW   ds; gene; cytostatic; antidiabetic; anorectic; cerebroprotective;
KW   neuroprotective; antiinflammatory; gene therapy; antisense therapy;
KW   thyromimetic; NOVX; pathology; cancer; diabetes; obesity;
KW   endocrine disorder; CNS disorder; inflammatory disorder;
KW   chromosome mapping; tissue typing; predictive medicine.
XX
OS   Homo sapiens.
XX
PN   WO2003000842-A2.
XX
PD   03-JAN-2003.
XX
PF   04-JUN-2002; 2002WO-US017443.
XX
PR   04-JUN-2001; 2001US-0295607P.
PR   04-JUN-2001; 2001US-0295661P.
PR   06-JUN-2001; 2001US-0296404P.
PR   06-JUN-2001; 2001US-0296418P.
PR   07-JUN-2001; 2001US-0296575P.
PR   11-JUN-2001; 2001US-0297414P.
PR   12-JUN-2001; 2001US-0295573P.
PR   12-JUN-2001; 2001US-0297567P.
PR   14-JUN-2001; 2001US-0298285P.
PR   15-JUN-2001; 2001US-0298528P.
PR   18-JUN-2001; 2001US-0299133P.
PR   19-JUN-2001; 2001US-0299230P.
PR   21-JUN-2001; 2001US-0299949P.
PR   22-JUN-2001; 2001US-0300177P.
PR   26-JUN-2001; 2001US-0300883P.
PR   28-JUN-2001; 2001US-0301530P.
PR   28-JUN-2001; 2001US-0301550P.
PR   03-JUL-2001; 2001US-0302951P.
PR   31-JUL-2001; 2001US-0308890P.
PR   14-SEP-2001; 2001US-0322297P.
PR   25-SEP-2001; 2001US-0324669P.
PR   03-DEC-2001; 2001US-0337477P.
PR   14-DEC-2001; 2001US-0341562P.
PR   21-FEB-2002; 2002US-0358656P.
PR   21-FEB-2002; 2002US-0359122P.
PR   22-FEB-2002; 2002US-0358978P.
PR   22-FEB-2002; 2002US-0359034P.
PR   22-FEB-2002; 2002US-0359035P.
PR   22-FEB-2002; 2002US-0359121P.
PR   27-FEB-2002; 2002US-0359964P.
PR   01-MAR-2002; 2002US-0360858P.
PR   12-MAR-2002; 2002US-0363430P.
PR   12-MAR-2002; 2002US-0363676P.
PR   10-APR-2002; 2002US-0371346P.
PR   10-MAY-2002; 2002US-0379444P.
PR   04-JUN-2002; 2002US-00379444.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML,  Anderson DW,  Berghs C,  Casman SJ,  Catterton E;
PI   Dipippo VA,  Edinger SR,  Eisen A,  Ellerman K,  Gangolli EA;
PI   Gerlach VL,  Gorman L,  Guo X,  Herrmann JL,  Hjalt T,  Ji W,  Kekuda R;
PI   Khramtsov NV,  Li L,  Liu X,  Malyankar UM,  Miller CE,  Millet I;
PI   Ort T,  Padigaru M,  Patturajan M,  Pena CEA,  Rastelli L,  Rieger DK;
PI   Rothenberg ME,  Shenoy SG,  Shimkets RA,  Smithson G,  Spaderna SK;
PI   Spytek KA,  Stone DJ,  Vernet CAM,  Zhong H,  Zhong M,  Alsobrook JP;
PI   Burgess CE,  Lepley DM;
XX
DR   WPI; 2003-210149/20.
DR   P-PSDB; ADC10082.
XX
PT   New isolated NOVX polypeptides and nucleic acid molecules useful for
PT   treating, preventing and diagnosing pathological conditions with NOVX-
PT   associated disorders, such as cancer, obesity, diabetes and inflammatory
PT   or CNS diseases.
XX
PS   Claim 20; SEQ ID NO 101; 772pp; English.
```

```
XX
CC   The invention relates to novel isolated polypeptides, mature form of the
CC   polypeptide, a sequence that is 95% identical to the polypeptide or the
CC   polypeptide comprising one or more conservative substitutions. The NOVX
CC   polypeptide is useful for treating or preventing a pathology associated
CC   with the polypeptide e.g. disorders associated with aberrant expression
CC   or activity of the polypeptide, such as cancer, diabetes, obesity, and
CC   endocrine, CNS and inflammatory disorders. They can also be used in
CC   various detection and screening assays, chromosome mapping, tissue typing
CC   and predictive medicine. This sequence corresponds to the coding sequence
CC   for one of the polypeptides of the invention.
XX
SQ   Sequence 2006 BP; 637 A; 390 C; 493 G; 486 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34.4;  DB 9;  Length 2006;
  Best Local Similarity   48.9%;  Pred. No. 2.7;
  Matches   92;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        208 GTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAAC 267
              ||| ||    | || |||||  |    |    |  | ||   ||  | ||| |  | |  ||
Db        540 GTTCTTGATCTAGCAGGAGTGATACAAAACGTCATAGATCACGTGCCAGAGCCTCACTAC 599

Qy        268 TTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTAC 327
              | |    | | ||  |  | |   || ||  || | ||   |    |  | |  | |
Db        600 CTGGTTGCTGTGGATGAGTCTGTTCATACCCTGGAAGACATAGTCAAGTGTATCAGTAAA 659

Qy        328 GGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTAC 387
              || | |  | |  | |||||||||   |  || | || ||  |  | ||| || ||
Db        660 AATACTGGCCCTGGGAAAATCCAGAAAATACCCAGAGAAAATGCATACCTAACCAAGGAC 719

Qy        388 GGTACGCA 395
              |||||
Db        720 TTAACGCA 727


RESULT 12
ACD26496/c
ID   ACD26496 standard; DNA; 18540 BP.
XX
AC   ACD26496;
XX
DT   09-SEP-2003  (first entry)
XX
DE   Human tumour necrosis factor receptor related protein-1, TRP-1, DNA.
XX
KW   Human; ds; gene; TRP-1; inflammatory disorder; rheumatoid arthritis;
KW   tumour necrosis factor receptor related protein-1; Crohn's disease; AIDS;
KW   asthma; irritable bowel syndrome; ulcerative colitis; multiple sclerosis;
KW   diabetes mellitus; immune disorder; autoimmune disease; allergy; stroke;
KW   systemic lupus erythematosus; neurodegenerative disorder; prostatitis;
KW   Alzheimer's disease; Parkinson's disease; cell proliferative disorder;
KW   cell differentiation disorder; ischaemic injury; myocardial infarction;
KW   apoptotic disorder; gastrointestinal and reproductive tract disorder;
KW   cancer; ulcer; bone disorder; osteoporosis; blood disorder; hepatitis;
KW   aplastic anaemia; thrombocytopenia; viral disorder; viral encephalitis;
KW   chromosome Xq12.
XX
OS   Homo sapiens.
XX
PN   US2003059889-A1.
XX
PD   27-MAR-2003.
XX
PF   28-AUG-2002; 2002US-00231426.
XX
PR   05-SEP-2001; 2001US-0317151P.
XX
PA   (BODI/) BODIAN D L.
PA   (LABO/) LABOW M A.
PA   (MICK/) MICKANIN C S.
XX
PI   Bodian DL,  Labow MA,  Mickanin CS;
XX
DR   WPI; 2003-521912/49.
DR   P-PSDB; ABU62739.
XX
PT   New tumor necrosis factor receptor related protein-1 polypeptide and
PT   nucleic acid, useful for diagnosing and treating disorders in a human,
PT   e.g. an inflammatory disorder, an immune disorder or a neurodegenerative
PT   disorder.
XX
PS   Claim 16; Page 21-30; 36pp; English.
XX
CC   The invention relates to a new isolated tumour necrosis factor receptor
CC   related protein-1 polypeptide, TRP-1. The tumour necrosis factor receptor
CC   related protein-1 polypeptide, nucleic acid and antibody are useful for
CC   diagnosing and treating disorders in a human, such as an inflammatory
CC   disorder (e.g. rheumatoid arthritis, Crohn's disease, AIDS, allergies,
CC   asthma, irritable bowel syndrome, ulcerative colitis, multiple sclerosis
CC   or diabetes mellitus), an immune disorder (e.g. autoimmune disease or
CC   systemic lupus erythematosus), a neurodegenerative disorder (e.g.
CC   Alzheimer's disease or Parkinson's disease), a cell proliferative
CC   disorder (e.g. various types of cancer), a cell differentiation disorder
CC   (e.g. stroke, ischaemic injury or myocardial infarction), an apoptotic
CC   disorder, a gastrointestinal and reproductive tract disorder (e.g. ulcers
CC   or prostatitis), bone disorder (e.g. osteoporosis), blood disorder (e.g.
CC   aplastic anaemia or thrombocytopenia) or viral disorder (e.g. hepatitis
CC   or viral encephalitis). The polypeptides are also useful in manufacturing
CC   a medicament for treating the disorders. The methods are useful for
CC   identifying molecules that bind to and/or modulate the activity of the
CC   polypeptide, or the transcription or translation of the nucleic acid. The
CC   present sequence represents human tumour necrosis factor receptor related
CC   protein-1, TRP-1, DNA which is located on chromosome Xq12
XX
SQ   Sequence 18540 BP; 5173 A; 3532 C; 3591 G; 6244 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34.4;  DB 8;  Length 18540;
  Best Local Similarity   52.0%;  Pred. No. 6.5;
  Matches   77;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         86 CAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTA 145
              ||||  |||||  |     | |  |    || || | |  | |   | ||  |   |||
Db       5513 CAGATGGTAATTATTAAAAATAAAATGACAGAAATTTTAAAAATTCACTGGATGGGCTTA 5454

Qy        146 TTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGT 205
              | ||  ||  |  || ||| ||||   ||| | ||   ||||  || ||| |  |||
Db       5453 GTAGTAGAGAAGAGACAGATGATAGAATCAGTGAACTTGAGGACAGAACAACAGAATTTA 5394

Qy        206 CCGTTATTTCACAAGAAGGAGGAAATAA 233
                  |  |||| | |||| |||| |
Db       5393 TGCAATATGAACAACAGGGAGAAAATGA 5366


RESULT 13
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
```

```
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34.2;  DB 3;  Length 456;
  Best Local Similarity   56.8%;  Pred. No. 1.8;
  Matches   63;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        275 ATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATA 334
              |||| |  ||  | ||  ||| |||| |||  | ||   ||  | ||  |   || ||
Db        206 AAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTA 265

Qy        335 GTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
                |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        266 CTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316


RESULT 14
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34;  DB 2;  Length 361;
  Best Local Similarity   56.1%;  Pred. No. 1.9;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
              ||   |||   || |  ||  |||  ||||  |||  | ||   ||  | ||  |   ||
Db        140 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 199

Qy        332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
                || |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        200 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 253


RESULT 15
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
```

```
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 34;  DB 2;  Length 361;
  Best Local Similarity   56.1%;  Pred. No. 1.9;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
              ||   |||   || |  ||  |||  ||||  |||  | ||   ||  | ||  |   ||
Db        140 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 199

Qy        332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
               || |    |  | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        200 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 253


Search completed: March 18, 2004, 09:11:47
Job time : 290.405 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	41	9.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	34	7.5	361	1	US-08-233-788A-56	Sequence 56, Appl
	3	34	7.5	456	1	US-08-233-788A-58	Sequence 58, Appl
	4	32.2	7.1	1443	4	US-09-328-352-2823	Sequence 2823, Ap
c	5	32.2	7.1	1830121	4	US-09-557-884-1	Sequence 1, Appli
c	6	32.2	7.1	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c	7	32	7.0	651	4	US-09-134-000C-1323	Sequence 1323, Ap
	8	31.8	7.0	1830121	4	US-09-557-884-1	Sequence 1, Appli
	9	31.8	7.0	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c	10	31.4	6.9	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	11	31	6.8	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	12	30.6	6.7	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	13	30.6	6.7	1350	4	US-09-328-352-3386	Sequence 3386, Ap
c	14	30.2	6.6	1053	4	US-09-543-681A-408	Sequence 408, App
	15	30.2	6.6	1072	4	US-08-956-171E-435	Sequence 435, App
	16	30	6.6	1454	2	US-08-657-392-1	Sequence 1, Appli
	17	30	6.6	1454	5	PCT-US94-02539-1	Sequence 1, Appli
	18	30	6.6	1455	2	US-08-657-392-26	Sequence 26, Appl
	19	30	6.6	1455	5	PCT-US94-02539-26	Sequence 26, Appl
	20	30	6.6	9072	4	US-08-956-171E-45	Sequence 45, Appl
	21	30	6.6	92407	4	US-09-596-002-36	Sequence 36, Appl
	22	29.8	6.5	640681	4	US-09-790-988-1	Sequence 1, Appli
c	23	29.6	6.5	474	4	US-09-621-976-18033	Sequence 18033, A
c	24	29.6	6.5	553	4	US-09-621-976-15491	Sequence 15491, A
	25	29.6	6.5	1588	1	US-08-446-908-3	Sequence 3, Appli
	26	29.6	6.5	1588	1	US-08-231-205A-3	Sequence 3, Appli
	27	29.6	6.5	1588	2	US-08-871-161-3	Sequence 3, Appli
	28	29.6	6.5	1589	4	US-09-023-655-1132	Sequence 1132, Ap
c	29	29.4	6.4	141	4	US-08-956-171E-2326	Sequence 2326, Ap
c	30	29.4	6.4	501	4	US-08-956-171E-654	Sequence 654, App
	31	29.4	6.4	1440	4	US-09-328-352-3130	Sequence 3130, Ap
	32	29.4	6.4	1710	4	US-09-543-681A-934	Sequence 934, App
	33	29.4	6.4	7253	4	US-09-268-347-35	Sequence 35, Appl
	34	29.4	6.4	9793	1	US-08-470-202-56	Sequence 56, Appl
	35	29.4	6.4	9793	1	US-08-471-770-56	Sequence 56, Appl
	36	29.4	6.4	9793	2	US-08-468-059-56	Sequence 56, Appl
	37	29.4	6.4	9793	3	US-09-109-916-56	Sequence 56, Appl
	38	29.4	6.4	9793	4	US-09-886-156-56	Sequence 56, Appl
	39	29.4	6.4	9793	4	US-09-886-149-56	Sequence 56, Appl
	40	29.4	6.4	9793	4	US-09-886-150-56	Sequence 56, Appl
	41	29.4	6.4	9793	4	US-09-886-159-56	Sequence 56, Appl
	42	29.2	6.4	1234	4	US-09-620-312D-953	Sequence 953, App
c	43	29.2	6.4	20284	4	US-09-526-193A-21	Sequence 21, Appl
	44	29	6.4	1287	4	US-09-461-325-18	Sequence 18, Appl
	45	29	6.4	1287	4	US-10-012-542-18	Sequence 18, Appl

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             9.0%;  Score 41;  DB 1;  Length 7218;
```

Best Local Similarity 3.0%; Pred. No. 0.0018;
Matches 11; Conservative 200; Mismatches 150; Indels 0; Gaps 0;

Qy 63 GACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTC 122
Db 1454 GAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395

Qy 123 TTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACG 182
Db 1394 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1335

Qy 183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAA 242
Db 1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1275

Qy 243 AGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGA 302
Db 1274 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1215

Qy 303 TGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGG 362
Db 1214 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155

Qy 363 AAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACA 422
Db 1154 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095

Qy 423 G 423
Db 1094 R 1094

RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 7.5%; Score 34; DB 1; Length 361;
Best Local Similarity 56.1%; Pred. No. 0.091;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
Db 140 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 199

Qy 332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
Db 200 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 253

RESULT 3
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 7.5%; Score 34; DB 1; Length 456;
Best Local Similarity 56.1%; Pred. No. 0.1;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331

```
Db        203 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 262

Qy        332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              || |   |   | |  ||||| ||||   ||||||| ||||  ||   |||||
Db        263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGT 316


RESULT 4
US-09-328-352-2823
; Sequence 2823, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2823
;   LENGTH: 1443
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-2823

  Query Match             7.1%;  Score 32.2;  DB 4;  Length 1443;
  Best Local Similarity   46.9%;  Pred. No. 0.74;
  Matches  100;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||   ||| ||| | |||  |  | |     | |       |   |||| || ||  ||
Db        481 ATATTAAAAAAACTATTAAATCAGTTACATGCTAATAATTATTTCCCTGTCATGGCTGCC 540

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              | |  | | | | ||||        |  |  || | |  | | ||  |  |  || | |
Db        541 GAGCTAGAGTTTTATCTTTTCTCACCTCAGCATCAATCCGAGATATGTTCAGAAAACCAA 600

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              | ||    ||   | ||| | | || |  |||    |    |||| |   ||  | || |
Db        601 TGTTTTGATATAGATGCGCCAAATAATTATCAGCAAGTTCTGGATGAAGTTGAAAAAGCA 660

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATT 213
                ||    |  |||||    | |  ||| ||||
Db        661 GCCCTACTACAATCAATTGAAATTACCGCTATT 693


RESULT 5
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
;   GENERAL INFORMATION:
;       APPLICANT: Fleischmann et al.
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/557,884
;            FILING DATE: 25-Apr-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/476,102
;            FILING DATE: JUN-5-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Michelle S. Marks
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB186P3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-309-8504
;            TELEFAX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             7.1%;  Score 32.2;  DB 4;  Length 1830121;
  Best Local Similarity   46.3%;  Pred. No. 25;
  Matches  106;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy          7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
              | ||| |  |  |  | |||   || |||| |||   |       ||   |||    |
Db     385834 ATCAACTATTGTTAAAGGATTCGGAAAATAATGGACTCTTTCTCGACATCAATTTGGGTT 385775

Qy         67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
              ||| |||     | |||||   |  ||  |  | | |  | | |    |  | ||  |||
Db     385774 AATGATGTAACTGTTCGTTACAACAATGGTCATACTGCTATTCACAATATGACGTTCTCA 385715

Qy        127 TTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAG 186
              | |||    |   | |||  ||    ||| |||  | ||   |||| ||
Db     385714 CTCAATAGTGGAACTATTTGTGCGTTAGTTGGCGTGAATGGAAGTGGCAAGTCCACATTA 385655

Qy        187 GAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATC 235
                |   | | |   |  || |   |||| |||| | || ||| |
Db     385654 TTTAAAAGCATAATGGGGCTTGTAAAGCCACAACAAGGCGAAATTAAAC 385606


RESULT 6
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;       APPLICANT: Robert D. Fleischmann
;                  Mark D. Adams
;                  Owen White
;                  Hamilton O. Smith
;                  J. Craig Venter
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville,
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/643,990A
;            FILING DATE: 23-Aug-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/487,429
;            FILING DATE: 1995-06-07
;            APPLICATION NUMBER: 08/426,787
```

```
;            FILING DATE: 1995-04-21
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Kenley K. Hoover
;            REGISTRATION NUMBER: 40,302
;            REFERENCE/DOCKET NUMBER: PB186P1C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-610-5790
;            TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match             7.1%;  Score 32.2;  DB 4;  Length 1830121;
  Best Local Similarity   46.3%;  Pred. No. 25;
  Matches  106;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy          7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
              | ||| |  |  |  | |||   || |||| |||   |       ||   |||    |
Db     385834 ATCAACTATTGTTAAAGGATTCGGAAAATAATGGACTCTTTCTCGACATCAATTTGGGTT 385775

Qy         67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
              |||  |||     | |||||   |  ||  |  | | |  | | |    |  | ||  |||
Db     385774 AATGATGTAACTGTTCGTTACAACAATGGTCATACTGCTATTCACAATATGACGTTCTCA 385715

Qy        127 TTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAG 186
              | |||    |   | |||  ||     ||| |||  | ||   |||| ||
Db     385714 CTCAATAGTGGAACTATTTGTGCGTTAGTTGGCGTGAATGGAAGTGGCAAGTCCACATTA 385655

Qy        187 GAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATC 235
                  | | |    |   || |  |||||  |||| | || ||| |
Db     385654 TTTAAAAGCATAATGGGGCTTGTAAAGCCACAACAAGGCGAAATTAAAC 385606


RESULT 7
US-09-134-000C-1323/c
; Sequence 1323, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1323
;   LENGTH: 651
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-1323

  Query Match             7.0%;  Score 32;  DB 4;  Length 651;
  Best Local Similarity   55.3%;  Pred. No. 0.58;
  Matches   83;  Conservative    0;  Mismatches   65;  Indels    2;  Gaps    1;

Qy        217 CAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTAT 276
              ||||||  | || | || |  |  |||  |  || || |||      |||     |  | |
Db        178 CAAGAACTACGATAAAAGCCAGTAAAATACTCCCTGGTAGGCCTAAATAGTAATTGGCAA 119

Qy        277 ATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGT 336
              ||| || ||  |||||| ||  || |  || |  | |||||  |  |   | || ||
Db        118 TTTGGGCTAA--GGCAATCCCGCCGTTTTCACTGATTGGAAAAGCCCCGATACTCATTGT 61

Qy        337 GCAGCTATTATCCAGAAAGGTTCTGGAAAT 366
                |  ||| | | || |||||   |  |||
Db         60 TCCCATATAAGCAAGTAAGGTATAGATAAT 31


RESULT 8
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;  GENERAL INFORMATION:
;       APPLICANT: Fleischmann et al.
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/557,884
;            FILING DATE: 25-Apr-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/476,102
;            FILING DATE: JUN-5-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Michelle S. Marks
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB186P3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-309-8504
;            TELEFAX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             7.0%;  Score 31.8;  DB 4;  Length 1830121;
  Best Local Similarity   52.7%;  Pred. No. 34;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
              ||||  |||| | ||||| |||| |  ||| | |||   |  ||  |  ||  ||  |||
Db     617140 AACACCTTGTAAATTATGGTGTTCATCATAATAGGTTGGAATGCCACCGCCCCCCGCACA 617199

Qy         67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
              ||||| |||   ||   ||   |  |    ||||       |    | || |  ||  | |
Db     617200 AATTACGATACTGCCTTTTTCAAGCAACCATTTGACCGGACGAATTTCAAAAATACGTTT 617259

Qy        127 TTTAATCAGGC 137
              |   |   |||
Db     617260 TGGTAATGGGC 617270


RESULT 9
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;       APPLICANT: Robert D. Fleischmann
;                  Mark D. Adams
;                  Owen White
```

```
;                 Hamilton O. Smith
;                 J. Craig Venter
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville,
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/643,990A
;            FILING DATE: 23-Aug-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/487,429
;            FILING DATE: 1995-06-07
;            APPLICATION NUMBER: 08/426,787
;            FILING DATE: 1995-04-21
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Kenley K. Hoover
;            REGISTRATION NUMBER: 40,302
;            REFERENCE/DOCKET NUMBER: PB186P1C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-610-5790
;            TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match             7.0%;  Score 31.8;  DB 4;  Length 1830121;
  Best Local Similarity   52.7%;  Pred. No. 34;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
              ||||  |||| | ||||| |||| |  |||.| |||  |  ||  |  ||  ||   |||
Db     617140 AACACCTTGTAAATTATGGTGTTCATCATAATAGGTTGGAATGCCACCGCCCCCCGCACA 617199

Qy         67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
              ||||| |||   ||  ||   |  |   ||||     |   | ||| |  ||  ||
Db     617200 AATTACGATACTGCCTTTTTCAAGCAACCATTTGACCGGACGAATTTCAAAAATACGTTT 617259

Qy        127 TTTAATCAGGC 137
              |  |  |||
Db     617260 TGGTAATGGGC 617270


RESULT 10
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
;  LENGTH: 1230025
;  TYPE: DNA
;  ORGANISM: Chlamydia pneumoniae
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(15000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (15001)..(30000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (30001)..(45000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (45001)..(60000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (60001)..(75000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (75001)..(90000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (90001)..(105000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (105001)..(120000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (120001)..(135000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (135001)..(150000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (150001)..(165000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (165001)..(180000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (180001)..(195000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (195001)..(210000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (210001)..(225000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (225001)..(240000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (240001)..(255000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (255001)..(270000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (270001)..(285000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (285001)..(300000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (300001)..(315000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (315001)..(330000)
;  OTHER INFORMATION: n=a or c or g or t
;  NAME/KEY: misc_feature
;  LOCATION: (330001)..(345000)
```

; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (720001)..(735000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (735001)..(750000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (750001)..(765000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (765001)..(780000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (780001)..(795000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (795001)..(810000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (810001)..(825000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

Query Match 6.9%; Score 31.4; DB 4; Length 1230025;
Best Local Similarity 56.2%; Pred. No. 38;
Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy         68 ATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCAT 127
              ||  |  |||| | |  ||||| | ||||  | |  | |  |||| ||    | ||||  |
Db     739726 ATCAAAATCTTGATTTTTCAGGAAATAAAATAGGCGAAGTTGAAGTAGCAGACTCTTTGT 739667

Qy        128 TTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTG 172
              ||    . | || |  |     |||| |   || |||| ||| ||
Db     739666 TTGCAGATGAGGGCGATGGCCTTCAATTGATCAAGGATTATATTG 739622
```

RESULT 11
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA

; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (300001)..(315000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t

```
;   NAME/KEY: misc_feature
;   LOCATION: (720001)..(735000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (735001)..(750000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (750001)..(765000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (765001)..(780000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (780001)..(795000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (795001)..(810000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (810001)..(825000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (825001)..(840000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (840001)..(855000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (855001)..(870000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (870001)..(885000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (885001)..(900000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (900001)..(915000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature

  Query Match             6.8%;  Score 31;  DB 4;  Length 1230025;
  Best Local Similarity   56.3%;  Pred. No. 52;
  Matches   58;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        344 TTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAG 403
              | |||   | |  | | || | | |||  ||||| | |     | | |   | |||||
Db    1044632 TGATCAGAAGAAATGCAGGCAGTGAGGTAAATATGAATCGTAGAGACATGGTAATAACAG 1044691

Qy        404 CAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCAC 446
              | || |||||| |       | || | ||||| ||  ||||||
Db    1044692 CTGTCGTAGTGAATGCTATATTGCTTGTGGCTCTTTTCGTCAC 1044734


RESULT 12
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 832;
  Best Local Similarity   11.1%;  Pred. No. 1.9;
  Matches   21;  Conservative   92;  Mismatches   76;  Indels    0;  Gaps    0;

Qy         10 AAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACAAAT 69
              ::::::: :::::: :| :: ::|::  ::::  |:  |::: ::|::::   |: :
Db         40 WRRKKKKAWWKYWKTWTWWYWRYAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYM 99

Qy         70 TATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTT 129
               :: :| ||  :::  :  :::: : :: :  : |     : :|:: :     : |: :::|
Db        100 AMWTRTWTGYAYYRSMMYWWRYRCWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKT 159

Qy        130 AATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAA 189
              : | ::  :: :|  ::   :  :  | | :  |::: :   :|:  ::| :: :  |
Db        160 YYWAATRYWWMMCWTKRWRASWWYCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGA 219

Qy        190 GGATCAAAA 198
              :::::::::
Db        220 MSWKYMWRM 228


RESULT 13
US-09-328-352-3386/c
; Sequence 3386, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3386
;   LENGTH: 1350
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-3386

  Query Match             6.7%;  Score 30.6;  DB 4;  Length 1350;
  Best Local Similarity   65.2%;  Pred. No. 2.5;
  Matches   45;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        309 TATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAA 368
              |||||  |||||  ||   |   || |||   |||| |||| ||| ||| | ||||||
Db        586 TATATGAAAAAGCTGTTTGTGAGCTAATGCCAGTATTTTCCATTAAGCTTCGGAAAATAA 527

Qy        369 GGCCAATAT 377
              |||| | |
Db        526 TGCCACTTT 518


RESULT 14
US-09-543-681A-408/c
; Sequence 408, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 408
;   LENGTH: 1053
;   TYPE: DNA
```

```
;   ORGANISM: Proteus mirabilis
US-09-543-681A-408

  Query Match             6.6%;  Score 30.2;  DB 4;  Length 1053;
  Best Local Similarity   62.7%;  Pred. No. 3;
  Matches   47;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        322 GCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACC 381
              || | |  ||| | | ||    ||||| || || |     ||||||||||| |   |||
Db        366 GCCTTCATTAACATTACAAGAGTTATCAAGGAAAGGATCAGAAATAAGGCCTTTCCAACC 307

Qy        382 CAGTACGGTACGCAG 396
                |  |||||||| |
Db        306 AATAACGGTACGCGG 292


RESULT 15
US-08-956-171E-435
; Sequence 435, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/956,171E
;             FILING DATE: 20-Oct-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 435:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1072 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 435:
US-08-956-171E-435

  Query Match             6.6%;  Score 30.2;  DB 4;  Length 1072;
  Best Local Similarity   49.3%;  Pred. No. 3;
  Matches   68;  Conservative    3;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         98 TTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCG 157
              ||  |||||: ||   |: |||      |    || |:|||    | || | |  ||  |
Db          6 TTTTGGTAARGGAGAAAWTCAATATGAAACAGTATAARGCGTATTTAATCGATTTAGATG 65

Qy        158 GCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCAC 217
              ||||     |||  |  | ||         || ||| ||| || |||   || |  || |
Db         66 GCACAATGTATATGGGAACAGATGAGATTGATGGAGCAANACAATTCATCGATTATTTAA 125

Qy        218 AAGAAGGAGGAAATAATC 235
              | | |  ||| | |  ||
Db        126 ATGTAAAAGGCATTCCTC 143


Search completed: March 18, 2004, 20:19:39
Job time : 65.4971 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		35	7.7	738	15	US-10-369-435-3	Sequence 3, Appli
2	c	34.4	7.5	18540	12	US-10-231-416-3	Sequence 3, Appli
3	c	34.4	7.5	18540	14	US-10-231-426-3	Sequence 3, Appli
4		34	7.5	529	15	US-10-027-632-241642	Sequence 241642,
5	c	33.8	7.4	203654	9	US-09-820-905-3	Sequence 3, Appli
6	c	33.4	7.3	598	15	US-10-027-632-206661	Sequence 206661,
7	c	33.4	7.3	598	15	US-10-027-632-206662	Sequence 206662,
8	c	33.4	7.3	598	15	US-10-027-632-206663	Sequence 206663,
9	c	33.4	7.3	660	15	US-10-369-493-29323	Sequence 29323, A
10		33.2	7.3	535	12	US-10-449-857A-18	Sequence 18, Appl
11		33.2	7.3	1444	12	US-10-425-114-3093	Sequence 3093, Ap
12		32.8	7.2	2649	15	US-10-094-749-1500	Sequence 1500, Ap
13		32.4	7.1	1914	9	US-09-815-242-4167	Sequence 4167, Ap
14		32.4	7.1	1929	9	US-09-815-242-8045	Sequence 8045, Ap
15		32.2	7.1	1163020	15	US-10-398-221-10	Sequence 10, Appl
16	c	32.2	7.1	1830121	14	US-10-329-960-1	Sequence 1, Appli
17	c	32.2	7.1	1830121	15	US-10-329-670-1	Sequence 1, Appli
18		32.2	7.1	3011208	15	US-10-398-221-2058	Sequence 2058, Ap
19		32	7.0	849	15	US-10-027-632-130879	Sequence 130879,
20		32	7.0	849	15	US-10-027-632-130880	Sequence 130880,
21		32	7.0	849	15	US-10-027-632-130881	Sequence 130881,
22	c	32	7.0	1353	12	US-10-282-122A-21063	Sequence 21063, A
23	c	32	7.0	1356	9	US-09-815-242-6450	Sequence 6450, Ap
24		32	7.0	1539	12	US-10-282-122A-25267	Sequence 25267, A
25	c	32	7.0	2410	12	US-10-424-599-51912	Sequence 51912, A
26	c	32	7.0	7664	9	US-09-070-927A-375	Sequence 375, App
27	c	31.8	7.0	583	15	US-10-027-632-14367	Sequence 14367, A
28	c	31.8	7.0	933	9	US-09-815-242-6967	Sequence 6967, Ap
29	c	31.8	7.0	933	12	US-10-282-122A-22047	Sequence 22047, A
30		31.8	7.0	1365	9	US-09-815-242-4243	Sequence 4243, Ap
31		31.8	7.0	1365	9	US-09-815-242-8196	Sequence 8196, Ap
32	c	31.8	7.0	2209	9	US-09-925-301-494	Sequence 494, App
33	c	31.8	7.0	5599	12	US-10-363-616-89	Sequence 89, Appl
34		31.8	7.0	1830121	14	US-10-329-960-1	Sequence 1, Appli
35		31.8	7.0	1830121	15	US-10-329-670-1	Sequence 1, Appli
36		31.6	6.9	480	9	US-09-815-242-4053	Sequence 4053, Ap
37		31.6	6.9	480	12	US-10-282-122A-7419	Sequence 7419, Ap
38	c	31.6	6.9	1251	15	US-10-369-493-44921	Sequence 44921, A
39		31.6	6.9	465237	9	US-09-933-267A-1	Sequence 1, Appli
40	c	31.4	6.9	472	9	US-09-864-761-2780	Sequence 2780, Ap
41	c	31.4	6.9	664	15	US-10-027-632-135259	Sequence 135259,
42	c	31.4	6.9	664	15	US-10-027-632-135260	Sequence 135260,
43	c	31.4	6.9	664	15	US-10-027-632-152059	Sequence 152059,
44	c	31.4	6.9	808	15	US-10-027-632-160710	Sequence 160710,
45	c	31.4	6.9	808	15	US-10-027-632-160711	Sequence 160711,

ALIGNMENTS

RESULT 1
US-10-369-435-3
; Sequence 3, Application US/10369435
; Publication No. US20040002440A1
; GENERAL INFORMATION:
; APPLICANT: Mathews, Sarah
; APPLICANT: Timms, Peter
; TITLE OF INVENTION: No. US20040002440A1el Diagnostic Agents and Uses Therefor
; FILE REFERENCE: 10338-15US (2615070/VPA)
; CURRENT APPLICATION NUMBER: US/10/369,435
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: AU PQ9540/00
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/AU01/01021
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(738)
US-10-369-435-3

Query Match 7.7%; Score 35; DB 15; Length 738;
Best Local Similarity 57.9%; Pred. No. 1.9;
Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy        340 GCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAA 399
              ||| ||   |||| |  | | || | | |||  ||||| |  |     |  | |  | |
Db          1 GCTGTTGATCAGAGAAATGCAGGCAGTGAGGTAAATATGAATCGTAGAGACATGGTAATA 60

Qy        400 ACAGCAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCAC 446
              ||||| || |||||| |       | || | ||||| ||  ||||||
Db         61 ACAGCTGTCGTAGTGAATGCTATATTGCTTGTGGCTCTTTTCGTCAC 107
```

```
RESULT 2
US-10-231-416-3/c
; Sequence 3, Application US/10231416
; Publication No. US20040043610A1
; GENERAL INFORMATION:
;  APPLICANT: Bodian, Dale L.
;  APPLICANT:  Labow, Mark
;  APPLICANT:  Mickanin, Craig S.
;  TITLE OF INVENTION: Tumor Necrosis Factor Related Protein-1
;  TITLE OF INVENTION:  Gene and Protein
;  FILE REFERENCE: 4-32139A
;  CURRENT APPLICATION NUMBER: US/10/231,416
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/317,151
;  PRIOR FILING DATE: 2002-09-05
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 18540
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-231-416-3

  Query Match             7.5%;  Score 34.4;  DB 12;  Length 18540;
  Best Local Similarity   52.0%;  Pred. No. 16;
  Matches   77;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         86 CAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTA 145
              ||||   ||||| |     | | |    || || | |  | |   | || |   |||
Db       5513 CAGATGGTAATTATTAAAAATAAAATGACAGAAATTTTAAAAATTCACTGGATGGGCTTA 5454

Qy        146 TTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGT 205
              | ||  ||  |  || ||| ||||   ||| | ||   ||||  || ||| |  |||
Db       5453 GTAGTAGAGAAGAGACAGATGATAGAATCAGTGAACTTGAGGACAGAACAACAGAATTTA 5394

Qy        206 CCGTTATTTCACAAGAAGGAGGAAATAA 233
                  |  |||| | |||| |||| |
Db       5393 TGCAATATGAACAACAGGGAGAAAATGA 5366


RESULT 3
US-10-231-426-3/c
; Sequence 3, Application US/10231426
; Publication No. US20030059889A1
; GENERAL INFORMATION:
;  APPLICANT: Bodian, Dale L.
;  APPLICANT:  Labow, Mark
;  APPLICANT:  Mickanin, Craig S.
;  TITLE OF INVENTION: Tumor Necrosis Factor Related Protein-1
;  TITLE OF INVENTION:  Gene and Protein
;  FILE REFERENCE: 4-32139A
;  CURRENT APPLICATION NUMBER: US/10/231,426
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/317,151
;  PRIOR FILING DATE: 2002-09-05
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 18540
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-231-426-3

  Query Match             7.5%;  Score 34.4;  DB 14;  Length 18540;
  Best Local Similarity   52.0%;  Pred. No. 16;
  Matches   77;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         86 CAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTA 145
              ||||   ||||| |     | | |    || || | |  | |   | || |   |||
Db       5513 CAGATGGTAATTATTAAAAATAAAATGACAGAAATTTTAAAAATTCACTGGATGGGCTTA 5454

Qy        146 TTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGT 205
              | ||  ||  |  || ||| ||||   ||| | ||   ||||  || ||| |  |||
Db       5453 GTAGTAGAGAAGAGACAGATGATAGAATCAGTGAACTTGAGGACAGAACAACAGAATTTA 5394

Qy        206 CCGTTATTTCACAAGAAGGAGGAAATAA 233
                  |  |||| | |||| |||| |
Db       5393 TGCAATATGAACAACAGGGAGAAAATGA 5366


RESULT 4
US-10-027-632-241642
; Sequence 241642, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 241642
;   LENGTH: 529
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(529)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-241642

  Query Match             7.5%;  Score 34;  DB 15;  Length 529;
  Best Local Similarity   54.9%;  Pred. No. 3.4;
  Matches   67;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy        132 TCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGG 191
              | | | | ||| ||  ||   |      || | ||| |   |||   || | ||||  |
Db        132 TGATGGGCCCAGTAGAAGTGTGGATATGACAGGGAATGGGAGCAGTTCACTCGAGGATAG 191

Qy        192 ATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCA 251
              |||||  |  | |  ||  | ||  |||| | |||| |||||  | |   |||   || ||
Db        192 ATCAATAAAACTTACCCATCTTGAACAACAGGGAGAAAATAGACTGAAAAAAAAAGAGCA 251

Qy        252 GG 253
              |
Db        252 AG 253


RESULT 5
US-09-820-905-3/c
; Sequence 3, Application US/09820905
; Patent No. US20020142938A1
; GENERAL INFORMATION:
;  APPLICANT: Yan, Chunhua
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001199
;  CURRENT APPLICATION NUMBER: US/09/820,905
```

```
;  CURRENT FILING DATE:  2001-03-30
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 203654
;   TYPE: DNA
;   ORGANISM: HUMAN
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(203654)
;   OTHER INFORMATION: n = A,T,C or G
US-09-820-905-3

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 203654;
  Best Local Similarity   50.3%;  Pred. No. 86;
  Matches   83;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy          5 AAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGA 64
              | ||  | || || ||| |  |  |     |||||    |  ||||||    |||   |
Db      71389 AGAAATATTTTTTCTTTTTTGTAGTATGTTTACTGAAAACTACTGGGAACAAAACATGGT 71330

Qy         65 CAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTT 124
              ||| | |      ||  || | | ||| |||  |||    | ||| | || || | |     |
Db      71329 CAACTGAACTAATGCTTCCTCCTGAAAATATGTTTTTAGAAAAAGTATAAACCCACAAAT 71270

Qy        125 CATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATA 169
                |   | ||||   |||      |   | || |  ||  | ||||
Db      71269 ACTAAGAACAGGAAACCAGATGAGAGTATGTGGAAACATAGAATA 71225


RESULT 6
US-10-027-632-206661/c
; Sequence 206661, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 206661
;   LENGTH: 598
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-206661

  Query Match             7.3%;  Score 33.4;  DB 15;  Length 598;
  Best Local Similarity   51.7%;  Pred. No. 5.6;
  Matches   76;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        200 TATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGA 259
              |||| | | ||| |||| |   |     | || |||  | | |||        ||  |
Db        563 TATTATTCTTTAATTCATAGAGATCTTGTAAGAATTAGCCTAAAACAATTAAGATACATT 504

Qy        260 ATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAA 319
                ||  ||  | || | ||||  |||   |   |||||  ||  ||   || || | | ||
Db        503 TTTTAAATGTGGCATTTATTAGGCAGTTGATAAATGCTAATTATTATAGCATTTAGTAAT 444

Qy        320 GCGCTTACGGTAATAGTGCAGCTATTA 346
                ||  | |   ||  | || || | | |
Db        443 GCTTTAATAATAGCATTGGAGGTGTGA 417


RESULT 7
US-10-027-632-206662/c
; Sequence 206662, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 206662
;   LENGTH: 598
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-206662

  Query Match             7.3%;  Score 33.4;  DB 15;  Length 598;
  Best Local Similarity   51.7%;  Pred. No. 5.6;
  Matches   76;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        200 TATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGA 259
              |||| | | ||| |||| |   |     | || |||  | | |||        ||  |
Db        563 TATTATTCTTTAATTCATAGAGATCTTGTAAGAATTAGCCTAAAACAATTAAGATACATT 504

Qy        260 ATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAA 319
                ||  ||  | || | ||||  |||   |   |||||  ||  ||   || || | | ||
Db        503 TTTTAAATGTGGCATTTATTAGGCAGTTGATAAATGCTAATTATTATAGCATTTAGTAAT 444

Qy        320 GCGCTTACGGTAATAGTGCAGCTATTA 346
                ||  | |   ||  | || || | | |
Db        443 GCTTTAATAATAGCATTGGAGGTGTGA 417


RESULT 8
US-10-027-632-206663/c
; Sequence 206663, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
```

```
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 206663
;   LENGTH: 598
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-206663

  Query Match             7.3%;  Score 33.4;  DB 15;  Length 598;
  Best Local Similarity   51.7%;  Pred. No. 5.6;
  Matches   76;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        200 TATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGA 259
              |||| | | ||| |||| |   |    | || ||| | | |||        ||  |
Db        563 TATTATTCTTTAATTCATAGAGATCTTGTAAGAATTAGCCTAAAACAATTAAGATACATT 504

Qy        260 ATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAA 319
              ||  ||  | || | ||||  |||    |   |||| ||  ||   || || | | ||
Db        503 TTTTAAATGTGGCATTTATTAGGCAGTTGATAAATGCTAATTATTATAGCATTTAGTAAT 444

Qy        320 GCGCTTACGGTAATAGTGCAGCTATTA 346
              ||  | |   ||  | || || | | |
Db        443 GCTTTAATAATAGCATTGGAGGTGTGA 417


RESULT 9
US-10-369-493-29323/c
; Sequence 29323, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 29323
;   LENGTH: 660
;   TYPE: DNA
;   ORGANISM: Caenorhabditis elegans
US-10-369-493-29323

  Query Match             7.3%;  Score 33.4;  DB 15;  Length 660;
  Best Local Similarity   55.7%;  Pred. No. 5.9;
  Matches   64;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        250 CAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGT 309
              ||  | || ||||       ||     | |||   | ||| |  | ||||  | || || |
Db        339 CAATCCGGCAATTTGCCGATTTGCCAGATTTGCCGAACAGCAATTGCCGCCCATTCCTGA 280

Qy        310 ATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAA 364
              ||      |  ||   ||| ||||||    |   | |||||| | ||| |||||||
Db        279 ATTATTTAGTGCATCTACAGTAATATCTTATAGAGTATCCATACAGGCTCTGGAA 225


RESULT 10
US-10-449-857A-18
```

```
; Sequence 18, Application US/10449857A
; Publication No. US20040043931A1
; GENERAL INFORMATION:
;  APPLICANT: Hershberg, Robert M.
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Mohamath, Raodoh
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF INFLAMMATORY BOWEL DISEASE
;  FILE REFERENCE: 584C1
;  CURRENT APPLICATION NUMBER: US/10/449,857A
;  CURRENT FILING DATE:  2003-05-30
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 18
;   LENGTH: 535
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Unknown Bacterium
US-10-449-857A-18

  Query Match             7.3%;  Score 33.2;  DB 12;  Length 535;
  Best Local Similarity   51.3%;  Pred. No. 6.1;
  Matches   77;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        248 ACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCA 307
              ||||| | |  |||  | ||| |||  |  ||||      ||||| | ||| || |||  |
Db         14 ACCAGACTGCAAATCGTCACTATGCACACGTTGACTGCCCGGGCCACGCCGACTATGTGA 73

Qy        308 GTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATA 367
              |     |  ||  | |   | | | |  | |  ||||       || ||| || |
Db         74 AGAACATGGTAACTGGTGCAGCTCAGATGGACGGTGCAATCCTTGTTTGTGCTGCAACTG 133

Qy        368 AGGCCAATATTACCCAGTACGGTACGCAGA 397
              | |   |||  | |||   ||  ||| |
Db        134 ACGGTCCTATGCCTCAGACACGTGAGCACA 163


RESULT 11
US-10-425-114-3093
; Sequence 3093, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3093
;   LENGTH: 1444
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700240986_FLI
US-10-425-114-3093

  Query Match             7.3%;  Score 33.2;  DB 12;  Length 1444;
  Best Local Similarity   55.1%;  Pred. No. 10;
  Matches   65;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        249 CCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAG 308
              ||| | ||   || || ||    | ||  |  || ||   |      |  ||   | || |
Db        207 CCATGGAGTTGATAATCACCCACCTTACGTCAAGAAACAAGTGGTGGAAAAGTTTCCCTG 266

Qy        309 TATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAAT 366
```

```
                  | || | || |||  | ||  ||||| || || | ||||| ||| || || || |
Db        267 CTCAGCGTACAGTGCTCCCTGTTCTAGTGAAGATAATTTCCAGGAAGCTTGTGCAATT 324
```

RESULT 12
US-10-094-749-1500
; Sequence 1500, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1500
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1500

Query Match 7.2%; Score 32.8; DB 15; Length 2649;
Best Local Similarity 48.4%; Pred. No. 19;
Matches 91; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy        208 GTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAAC 267
              ||| ||   | || |||||  |   |    | | ||   ||  | ||  |  | |  ||
Db        807 GTTCTTGATCTAGCAGGAGTGATACAAAACGTCATAGATCACGTGCCAAAGCTTCACTAC 866

Qy        268 TTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTAC 327
              | |    | | ||  |  | |   || ||   || | ||   |   |   | |  | |
Db        867 CTGGTTGCTGTGGATGAGTCTGTTCATACCCTGGAAGACATAGTCAAGTGTATCAGTAAA 926

Qy        328 GGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTAC 387
              || | |  | |  | |||||||||    |  || | || ||  |  | ||| || ||
Db        927 AATACTGGCCCTGGGAAAATCCAGAAAATACCCAGAGAAAATGCATACCTAACCAAGGAC 986

Qy        388 GGTACGCA 395
              |||||
Db        987 TTAACGCA 994
```

RESULT 13
US-09-815-242-4167
; Sequence 4167, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4167
; LENGTH: 1914
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4167

Query Match 7.1%; Score 32.4; DB 9; Length 1914;
Best Local Similarity 51.4%; Pred. No. 21;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        263 ATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCG 322
              | || || |    |  | | |||||  |  | | ||||   |  |||||||  |
Db       1001 AAAATTTAGAAATTGGTTATCAAACTGCAATTACCAAACCTATCAGTATAGAGGTCTCTA 1060

Qy        323 CTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCC 382
              || | |||  |||   |||   |  || |||  ||||||   | ||   ||||
Db       1061 AAGGCGATCATATAGCAATCATTGGGCCAAATGGTATTGGAAAATCGACCTTAATTAAAA 1120

Qy        383 AGTACGGTACGCAGAAAACAGCAGTT 408
              |  ||  ||  ||| |||| ||
Db       1121 CTATTGCTAATCAACAAAAAGCACTT 1146
```

RESULT 14
US-09-815-242-8045
; Sequence 8045, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

```
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8045
;   LENGTH: 1929
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1929)
US-09-815-242-8045

  Query Match             7.1%;  Score 32.4;  DB 9;  Length 1929;
  Best Local Similarity   51.4%;  Pred. No. 21;
  Matches   75;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        263 ATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCG 322
              | || || |     |  | | ||||| |  | | ||||   |  ||||||| |
Db       1001 AAAATTTAGAAATTGGTTATCAAACTGCAATTACCAAACCTATCAGTATAGAGGTCTCTA 1060

Qy        323 CTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCC 382
                || | |||  |||   |||   |  || |||  ||||||   | ||   ||||
Db       1061 AAGGCGATCATATAGCAATCATTGGGCCAAATGGTATTGGAAAATCGACCTTAATTAAAA 1120

Qy        383 AGTACGGTACGCAGAAAACAGCAGTT 408
                | ||  ||  ||| |||| ||
Db       1121 CTATTGCTAATCAACAAAAAGCACTT 1146


RESULT 15
US-10-398-221-10
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;   LENGTH: 1163020
;   TYPE: DNA
;   ORGANISM: Listeria innocua
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(end)
;   OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

  Query Match             7.1%;  Score 32.2;  DB 15;  Length 1163020;
  Best Local Similarity   49.8%;  Pred. No. 6.2e+02;
  Matches  108;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        215 CACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGT 274
              |||||||   ||  || ||| | ||  || |||   |   ||  ||  ||   ||  |
Db    1136649 CACAAGACATAGAGAAGAATTGTGCTGAATTCGTTAACTTAGATAAAAATCTATTAACTG 1136708

Qy        275 ATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATA 334
              |||||||   ||||||         |  |||    || ||  || ||   | |  |  |||
Db    1136709 ATATTGAAGTAACGGGGTGATTGTAAAATGAAGATAGATGTCAGAGAATTAAATGAGATA 1136768

Qy        335 GTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGC 394
              ||  ||| |  ||  || || | |   |||| |     || |||     ||
Db    1136769 GT-TAGCGAGAATATAGTAATGCTTAAGAAAGAGTATGAAAATTTAAATATAGCAATGAA 1136827

Qy        395 AGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATAT 431
              |  |  | || ||  |  | ||||||    | || |
Db    1136828 CGCTACAACCAATTTCACAGTAGAAACTCAGGAATTT 1136864


Search completed: March 19, 2004, 03:48:35
Job time : 260.734 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-2	Sequence 2, Appli
2	332.8	73.0	456	23	US-09-543-407-4	Sequence 4, Appli
3	295	64.7	561	17	US-09-252-691-171	Sequence 171, App
4	295	64.7	561	17	US-09-252-691C-171	Sequence 171, App
5	295	64.7	561	51	US-10-417-886-171	Sequence 171, App
6	74.4	16.3	100	45	US-10-146-492B-80	Sequence 80, Appl
7	66.4	14.6	100	45	US-10-146-492B-82	Sequence 82, Appl
8	64.8	14.2	100	45	US-10-146-492B-81	Sequence 81, Appl
9	44.2	9.7	438	31	US-09-739-449-3383	Sequence 3383, Ap
10	44.2	9.7	438	33	US-09-803-110-3383	Sequence 3383, Ap
c 11	44.2	9.7	29058	21	US-09-514-000-342	Sequence 342, App
12	44.2	9.7	187892	31	US-09-739-449-208	Sequence 208, App
13	44.2	9.7	187892	33	US-09-803-110-208	Sequence 208, App
c 14	41	9.0	7218	8	US-08-466-194-14	Sequence 14, Appl
15	37.6	8.2	456	23	US-09-543-407-13	Sequence 13, Appl
c 16	37.6	8.2	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
17	36.6	8.0	500	55	US-10-764-432-554	Sequence 554, App
18	36	7.9	1230230	20	US-09-438-185-1	Sequence 1, Appli
19	36	7.9	1230230	20	US-09-438-185A-1	Sequence 1, Appli
20	35.8	7.9	39	23	US-09-543-407-51	Sequence 51, Appl
21	35	7.7	738	50	US-10-369-435-3	Sequence 3, Appli
22	35	7.7	105325	1	PCT-US03-38193-1574	Sequence 1574, Ap
23	35	7.7	105325	54	US-10-723-860-1574	Sequence 1574, Ap
c 24	34.6	7.6	1362	104	US-60-475-871-2480	Sequence 2480, Ap
25	34.4	7.5	2006	46	US-10-160-619-101	Sequence 101, App
c 26	34.4	7.5	18540	47	US-10-231-416-3	Sequence 3, Appli
c 27	34.4	7.5	18540	47	US-10-231-426-3	Sequence 3, Appli
c 28	34.4	7.5	20537	39	US-09-949-016-12887	Sequence 12887, A
c 29	34.4	7.5	20538	39	US-09-949-016-17052	Sequence 17052, A
c 30	34.4	7.5	432287	39	US-09-947-911-180	Sequence 180, App
31	34.4	7.5	3227788	39	US-09-947-911-51	Sequence 51, Appl
32	34.2	7.5	456	23	US-09-543-407-25	Sequence 25, Appl
c 33	34.2	7.5	12757	36	US-09-897-516-2547	Sequence 2547, Ap
c 34	34.2	7.5	12757	36	US-09-897-516-2548	Sequence 2548, Ap
c 35	34.2	7.5	12757	36	US-09-897-516-2549	Sequence 2549, Ap
c 36	34.2	7.5	12757	36	US-09-897-516-2550	Sequence 2550, Ap
37	34.2	7.5	12757	36	US-09-897-516-2551	Sequence 2551, Ap
38	34.2	7.5	12757	36	US-09-897-516-2552	Sequence 2552, Ap
39	34.2	7.5	12757	36	US-09-897-516-2553	Sequence 2553, Ap
40	34.2	7.5	12757	36	US-09-897-516-2554	Sequence 2554, Ap
41	34.2	7.5	12757	36	US-09-897-516-2555	Sequence 2555, Ap
42	34.2	7.5	12757	36	US-09-897-516-2556	Sequence 2556, Ap
c 43	34.2	7.5	12757	36	US-09-897-516A-2547	Sequence 2547, Ap
c 44	34.2	7.5	12757	36	US-09-897-516A-2548	Sequence 2548, Ap
c 45	34.2	7.5	12757	36	US-09-897-516A-2549	Sequence 2549, Ap

ALIGNMENTS

```
RESULT 1
US-09-543-407-2
; Sequence 2, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
```

```
US-09-543-407-2

  Query Match           100.0%;  Score 456;  DB 23;  Length 456;
  Best Local Similarity  100.0%;  Pred. No. 4.9e-130;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
```

```
RESULT 2
US-09-543-407-4
; Sequence 4, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-543-407-4

  Query Match           73.0%;  Score 332.8;  DB 23;  Length 456;
  Best Local Similarity  83.1%;  Pred. No. 7.2e-92;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||  ||  |||||||  | |||  ||||||||||||  || ||||||||||||||  || |||
Db         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
```

```
                ||||||||||||||||||| ||||| ||||||||||  || ||  ||||||||||      ||
Db        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              ||  | |||||| ||||  |||||||| |||| |||||||| ||   |||||| |||||||
Db        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| |||     ||||||||||||||||||
Db        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| ||||||||||||  |
Db        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456


RESULT 3
US-09-252-691-171
; Sequence 171, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-171

  Query Match             64.7%;  Score 295;  DB 17;  Length 561;
  Best Local Similarity   78.0%;  Pred. No. 3.9e-80;
  Matches  355;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||  ||| |||||||||| ||| |||||||||||||||||||| | |||| |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||      |||  ||| | ||  |||| |||||||||||||| |||||||||||||  ||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||  |||||||| ||||||||||||||||   ||  | |  |||| | |    |||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||| || || ||| |  |||||||||||| || || || || ||||| || |||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              || || || ||| |||||| |||||| |||| ||||| || || || ||||| | ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||| |||||   |||  ||||||  || || |  ||    ||||||||||||| ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || || | ||| ||||| || ||||| |||||||||||||||||||||||||||| |||| |
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTA 455
              |||||||| |||||||||||||| | ||||| ||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 4
US-09-252-691C-171
; Sequence 171, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS.
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-171

  Query Match             64.7%;  Score 295;  DB 17;  Length 561;
  Best Local Similarity   78.0%;  Pred. No. 3.9e-80;
  Matches  355;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              |||||||||||  ||| |||||||||| ||| |||||||||||||||||||| | |||| |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||      |||  ||| | ||  |||| |||||||||||||| |||||||||||||  ||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||  |||||||| ||||||||||||||||   ||  | |  |||| | |    |||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||| || || ||| |  |||||||||||| || || || || ||||| || |||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              || || || ||| |||||| |||||| |||| ||||| || || || ||||| | ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||| |||||   |||  ||||||  || || |  ||    ||||||||||||| ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || || | ||| ||||| || ||||| |||||||||||||||||||||||||||| |||| |
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTA 455
              |||||||| |||||||||||||| | ||||| ||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 5
US-10-417-886-171
; Sequence 171, Application US/10417886
; GENERAL INFORMATION:
;  APPLICANT: Keith G. Weinstock et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.135
;  CURRENT APPLICATION NUMBER: US/10/417,886
;  CURRENT FILING DATE:  2003-04-17
```

```
;  PRIOR APPLICATION NUMBER: US/09/252,691C
;  PRIOR FILING DATE: 1999-02-18
;  PRIOR APPLICATION NUMBER: US 60/094,145
;  PRIOR FILING DATE: 1998-07-24
;  PRIOR APPLICATION NUMBER: US 60/074,787
;  PRIOR FILING DATE: 1998-02-18
;  NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-10-417-886-171

  Query Match             64.7%;  Score 295;  DB 51;  Length 561;
  Best Local Similarity   78.0%;  Pred. No. 3.9e-80;
  Matches  355;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||||||||  ||| ||||||||||| ||| ||||||||||||||||||| ||| || |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120
              ||     |||  ||| | ||  |||| ||||||||||||||| |||||||||||| ||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180
              ||||||  |||||||| ||||||||||||||    || | |  |||| | |    |||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240
              |||||||| || || ||| |  ||||||||||||| || || || || ||||| || |||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300
              || || || ||| ||||||  |||||| |||| ||||| || || || || ||||| | ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360
              ||||| ||||||   |||  ||||||  || || |  ||    |||||||||||||| ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420
              || || | ||| ||||| || ||||| ||||||||||||||||||||||||||||| |||| |
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTA 455
              |||||||| |||||||||||||| |  ||||| ||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 6
US-10-146-492B-80
; Sequence 80, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 80
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 39-58
US-10-146-492B-80

  Query Match             16.3%;  Score 74.4;  DB 45;  Length 100;
  Best Local Similarity   84.0%;  Pred. No. 4.7e-12;
  Matches   84;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        266 ACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTT 325
              || |||| |||||||| ||  |||||| |||||||||||||||||| |||||| | ||||
Db          1 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 60

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAA 365
              | ||||||| |||    |||||||||||||||||||| ||
Db         61 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 100


RESULT 7
US-10-146-492B-82
; Sequence 82, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 82
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 57-76
US-10-146-492B-82

  Query Match             14.6%;  Score 66.4;  DB 45;  Length 100;
  Best Local Similarity   79.0%;  Pred. No. 1.4e-09;
  Matches   79;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        123 TTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACG 182
              |||||||||||||| ||||| ||||||||||   || ||  ||||||||||     ||||
Db          1 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 60

Qy        183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAA 222
              |||| ||| |||||||| ||| | ||| || | ||||||
Db         61 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 100


RESULT 8
US-10-146-492B-81
; Sequence 81, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 81
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 38-57
US-10-146-492B-81

  Query Match             14.2%;  Score 64.8;  DB 45;  Length 100;
  Best Local Similarity   78.0%;  Pred. No. 4.4e-09;
  Matches   78;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;
```

```
Qy        208 GTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAAC 267
              ||| || | |||||||| | |  || ||||| ||  | |||||| |||| ||||||||
Db          1 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 60

Qy        268 TTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCA 307
              |||| |||||||| ||  |||||| ||||||||||||||
Db         61 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 100
```

RESULT 9
US-09-739-449-3383
; Sequence 3383, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 3383
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-3383

Query Match 9.7%; Score 44.2; DB 31; Length 438;
Best Local Similarity 47.6%; Pred. No. 0.017;
Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |      |||||||| |
Db         88 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 147

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
                |   |  ||  | || ||  | || |||      | | |||||   ||   | |
Db        148 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 207

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| || ||  | | |  ||    ||  ||| |||  | |      |   | || | ||
Db        208 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 267

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
               || | ||    || |  ||   ||||   || ||     | |   |||   | ||||||
Db        268 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 327

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
              |||| |     || |  || |  | || |||
Db        328 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 360
```

RESULT 10
US-09-803-110-3383
; Sequence 3383, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 3383
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-3383

Query Match 9.7%; Score 44.2; DB 33; Length 438;
Best Local Similarity 47.6%; Pred. No. 0.017;
Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |      |||||||| |
Db         88 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 147

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
                |   |  ||  | || ||  | || |||      | | |||||   ||   | |
Db        148 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 207

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| || ||  | | |  ||    ||  ||| |||  | |      |   | || | ||
Db        208 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 267

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
               || | ||    || |  ||   ||||   || ||     | |   |||   | ||||||
Db        268 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 327

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
              |||| |     || |  || |  | || |||
Db        328 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 360
```

RESULT 11
US-09-514-000-342/c
; Sequence 342, Application US/09514000
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)B
; CURRENT APPLICATION NUMBER: US/09/514,000
; CURRENT FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 15034
; SEQ ID NO 342
; LENGTH: 29058
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-514-000-342

Query Match 9.7%; Score 44.2; DB 21; Length 29058;
Best Local Similarity 47.6%; Pred. No. 0.073;
Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |      |||||||| |
Db      18682 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 18623

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
                |   |  ||  | || ||  | || |||      | | |||||   ||   | |
Db      18622 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 18563

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| || ||  | | |  ||    ||  ||| |||  | |      |   | || | ||
Db      18562 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 18503

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
               || | ||    || |  ||   ||||   || ||     | |   |||   | ||||||
Db      18502 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 18443

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
              |||| |     || |  || |  | || |||
Db      18442 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 18410
```

RESULT 12
US-09-739-449-208
; Sequence 208, Application US/09739449

```
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)C
;  CURRENT APPLICATION NUMBER: US/09/739,449
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 208
;   LENGTH: 187892
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(187892)
;   OTHER INFORMATION: unsure at all n locations
US-09-739-449-208

  Query Match             9.7%;  Score 44.2;  DB 31;  Length 187892;
  Best Local Similarity   47.6%;  Pred. No. 0.14;
  Matches  130;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |     ||||||| |
Db     137108 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 137167

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
                |  |  |  ||  | || ||  | || |||     | | |||||  ||   | |
Db     137168 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 137227

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| ||  ||  | | |  ||    ||  ||| |||  | |     |   | || | ||
Db     137228 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 137287

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
               || | ||    || |  ||   ||||   || ||     | |   |||   | ||||||
Db     137288 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 137347

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
               ||||  |    ||  |  ||  |  |  || |||
Db     137348 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 137380


RESULT 13
US-09-803-110-208
; Sequence 208, Application US/09803110
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)D
;  CURRENT APPLICATION NUMBER: US/09/803,110
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/739,449
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  PRIOR APPLICATION NUMBER: US 60/168,139
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 208
;   LENGTH: 187892
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(187892)
;   OTHER INFORMATION: unsure at all n locations
US-09-803-110-208

  Query Match             9.7%;  Score 44.2;  DB 33;  Length 187892;
  Best Local Similarity   47.6%;  Pred. No. 0.14;
  Matches  130;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        145 ATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTG 204
              || |  |||   |||  |   ||   |||| | |     ||||||| |
Db     137108 ATCGAACAATATGGCTGGTCCAACTCTGCCGGCGGCGCGCAGGAAGGGTATGGCAACCGA 137167

Qy        205 TCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
                |  |  |  ||  | || ||  | || |||     | | |||||  ||   | |
Db     137168 ATCCGCACCTACCAGAACGGCGGTTACAACCGGATCGTCGGCCACCAGTATGGTCGTCAC 137227

Qy        265 AACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCT 324
              ||| ||  ||  | | |  ||    ||  ||| |||  | |     |   | || | ||
Db     137228 AACCTTTCGGCTGTCGGTCAGGAAGGGCATGACAATTACGGTTCCACCACCCAGAACGGC 137287

Qy        325 TACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAG 384
               || | ||    || |  ||   ||||   || ||     | |   |||   | ||||||
Db     137288 AACCGCAACGTGGCCGGCATCGGCCAGTTCGGCTCCAACCACACCACCATCCTGACCCAG 137347

Qy        385 TACGGTACGCAGAAAACAGCAGTTGTAGTGCAG 417
               ||||  |    ||  |  ||  |  |  || |||
Db     137348 GACGGCAATGGCAACATTGCTGCAGGCGTTCAG 137380


RESULT 14
US-08-466-194-14/c
; Sequence 14, Application US/08466194
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, NW, Suite 500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20007
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,194
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/232,463
;      FILING DATE:  22-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/935,313
;      FILING DATE:  26-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/201 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
```

US-08-466-194-14

Query Match 9.0%; Score 41; DB 8; Length 7218;
Best Local Similarity 3.0%; Pred. No. 0.44;
Matches 11; Conservative 200; Mismatches 150; Indels 0; Gaps 0;

```
Qy        63 GACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTC 122
Db      1454 GAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395

Qy       123 TTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACG 182
Db      1394 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1335

Qy       183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAA 242
Db      1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1275

Qy       243 AGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGA 302
Db      1274 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1215

Qy       303 TGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGG 362
Db      1214 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155

Qy       363 AAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACA 422
Db      1154 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095

Qy       423 G 423
Db      1094 R 1094
```

RESULT 15
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match 8.2%; Score 37.6; DB 23; Length 456;
Best Local Similarity 51.2%; Pred. No. 2;
Matches 88; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy       272 CGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTA 331
Db       203 CGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATA 262

Qy       332 ATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTA 391
Db       263 GTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACG 322

Qy       392 CGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGGCTATTCGCGT 443
Db       323 CTAAAAACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGT 374
```

Search completed: March 19, 2004, 03:01:39
Job time : 2655.77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	36.6	8.0	500	1	PCT-US04-02000-554	Sequence 554, App
c	2	33.8	7.4	401433	6	US-10-765-790-79	Sequence 79, Appl
	3	32.8	7.2	418	6	US-10-779-543-9321	Sequence 9321, Ap
c	4	32.4	7.1	201	6	US-10-767-471-8836	Sequence 8836, Ap
c	5	32.4	7.1	201	6	US-10-767-471-34493	Sequence 34493, A
c	6	32.4	7.1	1227	6	US-10-767-471-501	Sequence 501, App
c	7	32.4	7.1	76573	6	US-10-767-471-10718	Sequence 10718, A
	8	32	7.0	1310	6	US-10-767-795-2887	Sequence 2887, Ap
c	9	31.8	7.0	5495	6	US-10-775-169-342	Sequence 342, App
c	10	31.2	6.8	201	6	US-10-767-471-8837	Sequence 8837, Ap
c	11	31.2	6.8	201	6	US-10-767-471-34439	Sequence 34439, A
	12	31	6.8	394468	7	US-60-548-091-5725	Sequence 5725, Ap
	13	30.8	6.8	309	6	US-10-257-047-43	Sequence 43, Appl
	14	30.8	6.8	309	6	US-10-257-047-46	Sequence 46, Appl
	15	30.8	6.8	3147	6	US-10-257-047-44	Sequence 44, Appl
	16	30.8	6.8	3147	6	US-10-257-047-45	Sequence 45, Appl
	17	30.8	6.8	3147	6	US-10-257-047-47	Sequence 47, Appl
	18	30.8	6.8	3147	6	US-10-257-047-48	Sequence 48, Appl
	19	29.8	6.5	592	6	US-10-767-701-25359	Sequence 25359, A
c	20	29.8	6.5	1035	6	US-10-779-543-7983	Sequence 7983, Ap
	21	29.8	6.5	2583	6	US-10-441-926-25	Sequence 25, Appl
	22	29.8	6.5	2583	6	US-10-441-926-27	Sequence 27, Appl
	23	29.8	6.5	2583	6	US-10-441-926-29	Sequence 29, Appl
	24	29.6	6.5	1084	6	US-10-487-578-48	Sequence 48, Appl
	25	29.6	6.5	1087	6	US-10-487-578-33	Sequence 33, Appl
	26	29.6	6.5	1217	6	US-10-487-578-46	Sequence 46, Appl
	27	29.6	6.5	1219	6	US-10-487-578-44	Sequence 44, Appl
	28	29.4	6.4	782	6	US-10-767-701-3762	Sequence 3762, Ap
c	29	29.4	6.4	188053	6	US-10-417-375A-125	Sequence 125, App
	30	29.2	6.4	344805	6	US-10-779-271-1	Sequence 1, Appli
	31	29.2	6.4	354592	6	US-10-765-790-70	Sequence 70, Appl
	32	29	6.4	1287	6	US-10-100-683-1253	Sequence 1253, Ap
c	33	29	6.4	1382	6	US-10-100-683-2753	Sequence 2753, Ap
c	34	29	6.4	113486	6	US-10-417-375A-18	Sequence 18, Appl
	35	29	6.4	156652	6	US-10-765-790-48	Sequence 48, Appl
	36	28.8	6.3	19959	6	US-10-767-471-10755	Sequence 10755, A
c	37	28.8	6.3	53118	6	US-10-767-471-10892	Sequence 10892, A
	38	28.6	6.3	346	6	US-10-779-543-12419	Sequence 12419, A
	39	28.6	6.3	675	5	US-09-741-873C-3	Sequence 3, Appli
	40	28.6	6.3	1743	6	US-10-417-884A-2080	Sequence 2080, Ap
c	41	28.6	6.3	3537	6	US-10-486-977-58	Sequence 58, Appl
	42	28.6	6.3	219715	6	US-10-417-375A-63	Sequence 63, Appl
	43	28.4	6.2	1149	6	US-10-771-241-139	Sequence 139, App
	44	28.4	6.2	2526	7	US-60-548-091-9	Sequence 9, Appli
	45	28.4	6.2	2665	7	US-60-548-091-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
PCT-US04-02000-554
; Sequence 554, Application PC/TUS0402000
; GENERAL INFORMATION:
; APPLICANT: Science Applications International Corporation
; TITLE OF INVENTION: Method and System for Identifying Biological Entities in
; TITLE OF INVENTION: Biological and Environmental Samples
; FILE REFERENCE: 36609-183264 (SAIC0087-PCT)
; CURRENT APPLICATION NUMBER: PCT/US04/02000
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US 60/441,745
; PRIOR FILING DATE: 2003-01-23
; PRIOR APPLICATION NUMBER: US 60/441,806
; PRIOR FILING DATE: 2003-01-23
; NUMBER OF SEQ ID NOS: 3242
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 554
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Clostridium perfringens
PCT-US04-02000-554

Query Match 8.0%; Score 36.6; DB 1; Length 500;
Best Local Similarity 44.4%; Pred. No. 0.053;
Matches 147; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

```
Qy         57 AACCGCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAG 116
              || | ||  || | | |||    |||||  | |   ||      |  ||||| |||  |
Db         68 AAACTCGTGAATTAAAGATGAAACTCGTGAAAAGGTTATGAAAGTTATAAATAAATATAA 127

Qy        117 CAAGTCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAG 176
              | |  || |   | |   ||    | | ||    ||       ||  ||| || ||
Db        128 CTATACTCCTAGTGCTATTGCTAGAAGTCTTTCAAAAAGTATTACTAATACTATTGGAGT 187

Qy        177 AGTACGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCG 236
              ||   |  ||     |||| | ||| |  |       | | |||||  || | ||
Db        188 TATAGTTCCTGATATTACAAATCCATTCTTTGGATCAATAATAAAAGGAATAAGTGATGT 247

Qy        237 GGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGC 296
              ||  ||||  | |             |   | ||  | ||  |||   |  |  |  |
Db        248 TGCAGAAGTGCATAATCTTAATTTGATTCTTTGTGACTCTAACGAGAGTATCGATAGGGA 307

Qy        297 CAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGG 356
              |   | || | || | |   ||  |   | | ||| | |  ||| ||      |
Db        308 AATAAAGGCTATTAAAACATTAAAAGAACAAAGAATTAGGGGAATTATAATTTGTCCAAC 367

Qy        357 TTCTGGAAATAAGGCCAATATTACCCAGTAC 387
              ||||   | || |     | || | |||||
Db        368 CTCTGTTGAAAATGATTTAAATAGCGAGTAC 398
```

```
RESULT 2
US-10-765-790-79/c
; Sequence 79, Application US/10765790
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79
;   LENGTH: 401433
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-79

  Query Match             7.4%;  Score 33.8;  DB 6;  Length 401433;
  Best Local Similarity   50.3%;  Pred. No. 2.6;
  Matches   83;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy          5 AAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGA 64
              | |||  | || || ||| |  |  |     |||||    |  ||||||    |||   |
Db     270713 AGAAATATTTTTTCTTTTTTGTAGTATGTTTACTGAAAACTACTGGGAACAAAACATGGT 270654

Qy         65 CAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTT 124
              ||| | |     ||  || | | ||| ||| ||| | ||| | || || | |    |
Db     270653 CAACTGAACTAATGCTTCCTCCTGAAAATATGTTTTTAGAAAAAGTATAAACCCACAAAT 270594

Qy        125 CATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATA 169
                |   | ||||  |||     |   | || |  ||  | ||||
Db     270593 ACTAAGAACAGGAAACCAGATGAGAGTATGTGGAAACATAGAATA 270549


RESULT 3
US-10-779-543-9321
; Sequence 9321, Application US/10779543
; GENERAL INFORMATION:
;  APPLICANT: Williams et al
;  TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
;  TITLE OF INVENTION:  IN CANCEROUS CELLS AND THEIR METHODS OF USE II
;  FILE REFERENCE: 2300-21302
;  CURRENT APPLICATION NUMBER: US/10/779,543
;  CURRENT FILING DATE:  2004-02-12
;  PRIOR APPLICATION NUMBER: 10/076,555
;  PRIOR FILING DATE: 2002-02-15
;  PRIOR APPLICATION NUMBER: 09/217,471
;  PRIOR FILING DATE: 1998-12-21
;  PRIOR APPLICATION NUMBER: 60/068,755
;  PRIOR FILING DATE: 1997-12-23
;  PRIOR APPLICATION NUMBER: 60/080,664
;  PRIOR FILING DATE: 1998-04-03
;  PRIOR APPLICATION NUMBER: 60/105,234
;  PRIOR FILING DATE: 1998-10-21
;  PRIOR APPLICATION NUMBER: 09/297,648
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: PCT/US99/01619
;  PRIOR FILING DATE: 1999-01-28
;  PRIOR APPLICATION NUMBER: 60/072,910
;  PRIOR FILING DATE: 1998-01-28
;  PRIOR APPLICATION NUMBER: 60/075,954
;  PRIOR FILING DATE: 1998-02-24
;  PRIOR APPLICATION NUMBER: 60/080,114
;  PRIOR FILING DATE: 1998-03-31
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 23767
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9321
;   LENGTH: 418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-779-543-9321

  Query Match             7.2%;  Score 32.8;  DB 6;  Length 418;
  Best Local Similarity   53.0%;  Pred. No. 0.74;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy         87 AGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTAT 146
              ||||| | |  || |    ||||   | |  | ||| || | | ||| | |  | |
Db        208 AGAATTTCAGACTGGGCGCCATGACCAAGGGCAATCTGCAGTCACTCAAGAGTACTTGGA 267

Qy        147 TGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGTC 206
              | |  | |  |||  | |  | |   |||| |  ||||||||||   | ||  | |
Db        268 AGATTGATTTAGCAGTGGTGCTTGGAACAGACTGGGCCAGGAAGGGCTATAAGCAGGGAG 327

Qy        207 CGTTATTTCACA 218
              || |||| | |
Db        328 AGTCATTTAAAA 339


RESULT 4
US-10-767-471-8836/c
; Sequence 8836, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8836
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-8836

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 201;
  Best Local Similarity   55.6%;  Pred. No. 0.8;
  Matches   60;  Conservative    1;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | || | ||  |  || |  |||||  | |||  ||  |||:|||    |||| |||||
Db        142 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 83

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |   ||   |   |  |  | || ||||| |||||    |  |||
Db         82 AGGATTATCATGTACATGATGAGGCAGAGCAGGAACAGAAAAAGCTGG 35


RESULT 5
US-10-767-471-34493/c
; Sequence 34493, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34493
;   LENGTH: 201
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-10-767-471-34493

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 201;
  Best Local Similarity   55.6%;  Pred. No. 0.8;
  Matches   60;  Conservative    1;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | ||| | ||  |  || |  |||||  | |||  ||  |||:|||    |||| |||||
Db        142 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 83

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |   ||   |  |  | | || |||| |||||    |  |||
Db         82 AGGATTATCATGTACATGATGAGGCAGAGCAGGAACAGAAAAAGCTGG 35


RESULT 6
US-10-767-471-501/c
; Sequence 501, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 501
;   LENGTH: 1227
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-501

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 1227;
  Best Local Similarity   55.6%;  Pred. No. 1.3;
  Matches   60;  Conservative    1;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | ||| | ||  |  || |  |||||  | |||  ||  |||:|||    |||| |||||
Db        476 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 417

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | |.|   ||   |  |  | | || |||| |||||    |  |||
Db        416 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 369


RESULT 7
US-10-767-471-10718/c
; Sequence 10718, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10718
;   LENGTH: 76573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(76573)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10718

  Query Match             7.1%;  Score 32.4;  DB 6;  Length 76573;
  Best Local Similarity   55.6%;  Pred. No. 4.4;
  Matches   60;  Conservative    1;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
              | ||| | ||  |  || |  |||||  | |||  ||  |||:|||    |||| |||||
Db      69821 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 69762

Qy        386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
              | | |   ||   |  |  | | || |||| |||||    |  |||
Db      69761 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 69714


RESULT 8
US-10-767-795-2887
; Sequence 2887, Application US/10767795
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Zhou, Yihua
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53534)B
;  CURRENT APPLICATION NUMBER: US/10/767,795
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 2887
;   LENGTH: 1310
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C71322_1
US-10-767-795-2887

  Query Match             7.0%;  Score 32;  DB 6;  Length 1310;
  Best Local Similarity   55.4%;  Pred. No. 1.8;
  Matches   62;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              |||   |  ||||||    |||| | ||  || | ||  ||    |||   || |   |
Db        626 TGATGCATCCATTGTTCCCAATGAGGATGGGAACATTGCCTGAAAATGGCTATGAATGGG 685

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATT 113
                | ||||||   | ||.||    ||| |  ||||   |||  ||||||| |
Db        686 TGTCAAATTCCAAGCTTGCCAACTCAAACGATAACAATGCTCAAAATGAAGT 737


RESULT 9
US-10-775-169-342/c
; Sequence 342, Application US/10775169
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Burczynski, Michael
;  APPLICANT:  Twine, Natalie
;  APPLICANT:  Dorner, Andrew
;  APPLICANT:  Trepicchio, William
;  TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
;  FILE REFERENCE: AM101080 (031896-013000)
;  CURRENT APPLICATION NUMBER: US/10/775,169
;  CURRENT FILING DATE:  2004-02-11
;  NUMBER OF SEQ ID NOS: 5278
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 342
;   LENGTH: 5495
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-169-342

  Query Match             7.0%;  Score 31.8;  DB 6;  Length 5495;
  Best Local Similarity   51.0%;  Pred. No. 3.2;
  Matches   75;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              || ||||||   ||||  |-    ||   | ||     |   |        || |   | |
Db       4776 TGCAAAACAGGATGTTTCTCCGTTTGAAAAAAAACAAGCAAGGAAAACAGAGTCATATCA 4717

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
```

```
Db          4716 CGTAAGCGCTTGTTCAGGGAAAATATGACCTTAAGTATGCTGTAAATGGATTTGAAAAGG 4657

Qy           122 CTTCATTTAATCAGGCGGCCATTATTG 148
                 |||||||||||||||||||||||||||
Db          4656 CTCCTTTGAGTTAGGCGGGCTTTTGTG 4630

RESULT 10
US-10-767-471-8837/c
; Sequence 8837, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8837
;  LENGTH: 201
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-767-471-8837

  Query Match
  Best Local Similarity   6.8%;  Score 31.2;  DB 6;  Length 201;
  Matches   60;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy          326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          183 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATGAGGAGGCTATTTCCCAGG 124

Qy          386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db          123 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 76

RESULT 11
US-10-767-471-34439/c
; Sequence 34439, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34439
;  LENGTH: 201
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-767-471-34439

  Query Match
  Best Local Similarity   6.8%;  Score 31.2;  DB 6;  Length 201;
  Matches   60;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy          326 ACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGT 385
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          183 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATGAGGAGGCTATTTCCCAGG 124

Qy          386 ACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACAGTCGCATATGG 433
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db          123 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 76

RESULT 12
US-60-548-091-5725
, Sequence 5725, Application US/60548091
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STROKE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001506
;  CURRENT APPLICATION NUMBER: US/60/548,091
;  CURRENT FILING DATE:  2004-02-27
;  NUMBER OF SEQ ID NOS: 24433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5725
;  LENGTH: 394468
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(394468)
;  OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables
US-60-548-091-5725

  Query Match
  Best Local Similarity   6.8%;  Score 31;  DB 7;  Length 394468;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0,

Qy          233 ATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCA 292
                ||| |||| ||||| |||| |||| |||| |||| |||||| | |||||| |||||||||
Db       320889 ATCTCTTGATAGTTGAGGAGGCTGTGAGTATACGCTTTGCTCACCAGGTGCAAACAGCTG 320948

Qy          293 ATGCCAACGATGCCAGTATATCGCAAAGCGCTTAC 327
                ||| |||| |||||| || |||| |||| |||||
Db       320949 GTGTCAATAACCCCAGTAAGTCCCCAAGTGTTTCC 320983

RESULT 13
US-10-257-047-43
; Sequence 43, Application US/10257047
; GENERAL INFORMATION:
;  APPLICANT: STEWARD, MICHAEL
;  APPLICANT:  COX, VIVIENNE FRANCES
;  TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
;  FILE REFERENCE: 37945-0041
;  CURRENT APPLICATION NUMBER: US/10/257,047
;  CURRENT FILING DATE:  2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/GB01/01599
;  PRIOR FILING DATE: 2001-04-09
;  PRIOR APPLICATION NUMBER: GB 0008582.9
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
;  LENGTH: 309
;  TYPE: DNA
;  ORGANISM: Plasmodium falciparum
US-10-257-047-43

  Query Match
  Best Local Similarity   6.8%;  Score 30.8;  DB 6;  Length 309;
  Matches   65;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy            2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
                ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db          132 TGAGAACCCAAACCCTACCTGTAACGAGAACAACGGTGGATGCGACGCTGACGCTAAGTG 191

Qy           62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAAGT 121
                |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db          192 CACCGAAGAAGACTCTGGTTCTAACGGAAAGAAGATTACTTGCGAATGTACTAAGCCAGA 251

Qy          122 CT 123
                ||
Db          252 CT 253

RESULT 14
US-10-257-047-46
```

```
; Sequence 46, Application US/10257047
; GENERAL INFORMATION:
;  APPLICANT: STEWARD, MICHAEL
;  APPLICANT:  COX, VIVIENNE FRANCES
;  TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
;  FILE REFERENCE: 37945-0041
;  CURRENT APPLICATION NUMBER: US/10/257,047
;  CURRENT FILING DATE:  2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/GB01/01599
;  PRIOR FILING DATE: 2001-04-09
;  PRIOR APPLICATION NUMBER: GB 0008582.9
;  PRIOR FILING DATE: 2000-04-08
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
;   LENGTH: 309
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-10-257-047-46

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 309;
  Best Local Similarity   53.3%;  Pred. No. 2.8;
  Matches   65;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              ||| || | ||    ||  | | | |   ||||   ||| |||| |   ||    || |
Db        132 TGAGAACCCAAACCCTACCTGTAACGAGAACAACGGTGGATGCGACGCTGACGCTAAGTG 191

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
              |  | ||  |   ||||| ||   | |||   |  ||    |  |||| | ||||| ||
Db        192 CACCGAAGAAGACTCTGGTTCTAACGGAAAGAAGATTACTTGCGAATGTACTAAGCCAGA 251

Qy        122 CT 123
              ||
Db        252 CT 253


RESULT 15
US-10-257-047-44
; Sequence 44, Application US/10257047
; GENERAL INFORMATION:
;  APPLICANT: STEWARD, MICHAEL
;  APPLICANT:  COX, VIVIENNE FRANCES
;  TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
;  FILE REFERENCE: 37945-0041
;  CURRENT APPLICATION NUMBER: US/10/257,047
;  CURRENT FILING DATE:  2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/GB01/01599
;  PRIOR FILING DATE: 2001-04-09
;  PRIOR APPLICATION NUMBER: GB 0008582.9
;  PRIOR FILING DATE: 2000-04-08
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Vector
;   OTHER INFORMATION: pVK104-01
US-10-257-047-44

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 3147;
  Best Local Similarity   53.3%;  Pred. No. 5.5;
  Matches   65;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              ||| || | ||    ||  | | | |   ||||   ||| |||| |   ||    || |
Db        198 TGAGAACCCAAACCCTACCTGTAACGAGAACAACGGTGGATGCGACGCTGACGCTAAGTG 257

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
              |  | ||  |   ||||| ||   | |||   |  ||    |  |||| | ||||| ||
Db        258 CACCGAAGAAGACTCTGGTTCTAACGGAAAGAAGATTACTTGCGAATGTACTAAGCCAGA 317

Qy        122 CT 123
              ||
Db        318 CT 319


Search completed: March 19, 2004, 03:09:57
Job time : 56.1445 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtcacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		132.2	29.0	220	12	BI941542	BI941542 dg26g06.y
2	c	53.6	11.8	207	9	AI368352	AI368352 SWOvMfCAR
3		39.4	8.6	1101	29	CNS00E93	AL068795 Drosophil
4	c	39	8.6	680	12	BJ618688	BJ618688 BJ618688
5		38.8	8.5	775	9	AI664697	AI664697 TENG0665
6	c	38	8.3	288	28	AZ578608	AZ578608 25d10 Sho
7		37.8	8.3	675	12	BP007561	BP007561 BP007561
8	c	37.8	8.3	715	9	AV896333	AV896333 AV896333
9		37.6	8.2	601	13	BU213938	BU213938 603108484
10	c	37.4	8.2	1214	28	CC310289	CC310289 TAM32-32B
11		36.8	8.1	708	13	BU387712	BU387712 603857374
12		36.8	8.1	924	28	AZ529672	AZ529672 ENTBM10TR
13	c	36.4	8.0	1550	28	CC211905	CC211905 CH261-186
14	c	36	7.9	499	28	BH528896	BH528896 BOGPO61TR
15		36	7.9	579	12	BG226083	BG226083 kq08h12.y
16	c	36	7.9	795	28	BZ229556	BZ229556 CH230-353
17	c	36	7.9	797	28	BH505143	BH505143 BOGBM31TF
18		36	7.9	975	29	CNS03TIU	AL259887 Tetraodon
19		35.8	7.9	883	28	BH159159	BH159159 ENTRB03TR
20		35.6	7.8	412	10	AW140314	AW140314 SWAMCAC45
21	c	35.6	7.8	760	28	BZ756658	BZ756658 PUFBN68TD
22		35.2	7.7	659	12	BJ481788	BJ481788 BJ481788
23	c	35.2	7.7	791	14	CB310770	CB310770 AGENCOURT
24		34.8	7.6	719	12	BG548059	BG548059 602575980
25		34.8	7.6	1201	13	BX406892	BX406892 BX406892
26		34.6	7.6	606	10	BE581316	BE581316 kq49c03.y
27	c	34.6	7.6	698	28	BH943841	BH943841 odf17b10.
28		34.6	7.6	700	13	BW245005	BW245005 BW245005
29		34.6	7.6	710	9	AV940831	AV940831 AV940831
30		34.4	7.5	435	14	CA856583	CA856583 PfESToac1
31		34.4	7.5	471	28	AQ055856	AQ055856 CIT-HSP-2
32		34.4	7.5	576	12	BM273661	BM273661 PfESToaa5
33		34.4	7.5	577	13	BQ451444	BQ451444 PfESToaa9
34		34.4	7.5	581	14	CB389390	CB389390 OSTF116E9
35		34.4	7.5	633	12	BJ104432	BJ104432 BJ104432
36	c	34.4	7.5	735	28	BZ110007	BZ110007 CH230-164
37		34.4	7.5	1094	29	CNS00BR5	AL057531 Drosophil
38	c	34.2	7.5	666	13	BU582719	BU582719 rd26b04.y
39	c	34.2	7.5	689	13	BQ625251	BQ625251 rd26b04.y
40	c	34.2	7.5	784	28	BZ280200	BZ280200 CH230-297
41		34.2	7.5	854	12	BM012028	BM012028 603636451
42		34.2	7.5	965	9	AL549572	AL549572 AL549572
43		34.2	7.5	1084	28	CC299970	CC299970 CH261-89H
44		34.2	7.5	1224	14	CF238215	CF238215 AGENCOURT
45		34	7.5	402	9	AW043435	AW043435 SWAMCAC37

ALIGNMENTS

RESULT 1
BI941542
LOCUS BI941542 220 bp mRNA linear EST 19-FEB-2003
DEFINITION dg26g06.y1 Xenopus laevis gastrula non normalized Xenopus laevis cDNA clone IMAGE:3749963 5' similar to SW:CSGB_ECOLI P39828 MINOR CURLIN SUBUNIT PRECURSOR. ;, mRNA sequence.
ACCESSION BI941542
VERSION BI941542.1 GI:16256014
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 220)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Source lab clone id - 3749963
Trace considered overall poor quality This clone is available royalty-free through LLNL; contact the IMAGE Consortium (image@image.llnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .220
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:3749963"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F'"
/clone_lib="Xenopus laevis gastrula non normalized"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2 ug of poly A+ RNA (equal parts from stage 10.5 and stage 11.5 gastrulae). EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. The library was mass excised and used to infect Top10F'. Clones were picked into freezing medium (per liter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM K2HPO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4 7 H2O, 6.8 mM (NH4)2SO4, 4 % w/v glycerol) and grown for 24 hours. Original library construction by Bruce Blumberg (Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

Query Match 29.0%; Score 132.2; DB 12; Length 220;
Best Local Similarity 75.6%; Pred. No. 5.2e-30;
Matches 164; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy         80 CTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGG 139
              ||  |||| |||||||  || |||| | ||||||  || ||||||||||||||||||| |
Db          3 CTAATTCATAATATAACTTCGCGGCACATGAATCGAGTAAGTCTTCATTTAATCAGGCAG 62

Qy        140 CCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAAC 199
              |||| ||||||||||  || ||  ||||||||||     |||  |||| ||| |||||||
Db         63 CCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACTGCAGGGAGGCTCAAAAC 122

Qy        200 TATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGA 259
              | ||| | |||  | | |||||||  | |  || ||||| ||  | |||||| ||||
Db        123 TTTTGGCGGTTGCTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAG 182

Qy        260 ATTATAACTTTGCGTATATTGAGCAAACGGGCAATGC 296
              |||||||| |||| |||||||| ||  |||||| | |
Db        183 ATTATAACCTTGCATATATTGATCAGGCGGGCAGTCC 219
```

RESULT 2
AI368352/c
LOCUS AI368352 207 bp mRNA linear EST 11-JAN-1999
DEFINITION SWOvMfCAR04G09SK Onchocerca volvulus microfilaria cDNA (SAW98MLW-OvMf) Onchocerca volvulus cDNA clone SWOvMfCAR04G09 5', mRNA sequence.
ACCESSION AI368352
VERSION AI368352.1 GI:4147105
KEYWORDS EST.
SOURCE Onchocerca volvulus
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 207)
AUTHORS Williams,S.A.
TITLE Genes expressed in microfilaria of Onchocerca volvulus
JOURNAL Unpublished (1999)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.

FEATURES Location/Qualifiers
source 1. .207
/organism="Onchocerca volvulus"
/mol_type="mRNA"
/db_xref="taxon:6282"
/clone="SWOvMfCAR04G09"
/dev_stage="microfilaria"
/lab_host="XL1-Blue MRF'"
/clone_lib="Onchocerca volvulus microfilaria cDNA (SAW98MLW-OvMf)"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.8 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email:genome@smith.edu."

ORIGIN

Query Match 11.8%; Score 53.6; DB 9; Length 207;
Best Local Similarity 91.8%; Pred. No. 1.4e-05;
Matches 56; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60
              ||||||| | |||||| |||||||||||| |||||||||||||||||||||||||||| ||
Db         61 ATGAAAACCNAATTGTCATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 2

Qy         61 G 61
              |
Db          1 G 1
```

RESULT 3
CNS00E93
LOCUS CNS00E93 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR28G06 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL068795
VERSION AL068795.1 GI:4949039
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR28G06"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             8.6%;  Score 39.4;  DB 29;  Length 1101;
  Best Local Similarity   33.2%;  Pred. No. 0.61;
  Matches   92;  Conservative   52;  Mismatches  133;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCG 61
              :|||||| :  ||| ||  : |||:  | : || |::|  |  : ||    |::     :
Db        825 KGAAAAAAWCCTTGATAAAYGTGAWAATWDAAAAASKGTTTCTSTCTTTTTTWDTTGGGS 884

Qy         62 CGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGT 121
               | : || |||     ||   : : | |::   : ::|:   |      ::|:|   |::
Db        885 TGCMTAAATATYKRGGGGAAGRGKTAAARDTSGDAKKTRTTATTBSATSMBTWAATTARB 944

Qy        122 CTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTAC 181
              | |:||   | :    || || || :|    |  :  :||  |::  |:   :  |
Db        945 KSTAAWTTKGSCCKAWCCCCTTTTTTTBTTTTTTTTKGGVGGGAAWWTTTKKTTKGTTTT 1004

Qy        182 GCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGA 241
              |:| |||   :|||   |||     :| |: | |::| |:||::|:  |  ||: ||
Db       1005 TKKARGGAGGGGGRAAAAATTTGGAATGWAATDAARAWRAGGRAGRRARAKAARGGRGGA 1064

Qy        242 AAGTCGACCAGGCAGGGAATTATAACTTTGCGTATAT 278
              ||   ||  ||  |:  : :  :   |  : | ::||
Db       1065 AAWRGGAAAAGATADAAWTKGTWTTTTGGRTGAWWAT 1101


RESULT 4
BJ618688/c
LOCUS       BJ618688                680 bp    mRNA    linear   EST 01-OCT-2003
DEFINITION  BJ618688 NIBB Mochii normalized Xenopus early gastrula library
            Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION   BJ618688
VERSION     BJ618688.1  GI:37256713
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
            Kohara,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp
            The information of this clone is available through the following
            URL.
            http://xenopus.nibb.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match             8.6%;  Score 39;  DB 12;  Length 680;
  Best Local Similarity   89.4%;  Pred. No. 0.69;
  Matches   42;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        410 TAGTGCAGAAACAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
              ||||||||| |||||||||| |||||||||||||| || ||||| |||
Db        680 TAGTGCAGAGACAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 634


RESULT 5
AI664697
LOCUS       AI664697                775 bp    mRNA    linear   EST 07-JUL-1999
DEFINITION  TENG0665 T. Cruzi epimastigote normalised cDNA Library Trypanosoma
            cruzi cDNA clone n1059.r 5', mRNA sequence.
ACCESSION   AI664697
VERSION     AI664697.1  GI:4775685
KEYWORDS    EST.
SOURCE      Trypanosoma cruzi
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   Gonzalez Rey,E., Remisz,E., Delgado Garcia,A. and Gonzalez,A.
  TITLE     Characterization of ESTs from Trypanosoma cruzi epimastigotes
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Delgado Alberto
            Departamento de Biologia Molecular, Lab 303
            Instituto de Parasitologia y Biomedicina
            Consejo Superior de Investigaciones Cientificas C/ Ventanilla No
            11, E-18001, Granada, Spain
            Tel: 34 958 805058
            Fax: 34 958 203323
            Seq primer: T7
            High quality sequence stop: 775.
FEATURES             Location/Qualifiers
     source          1. .775
                     /organism="Trypanosoma cruzi"
                     /mol_type="mRNA"
                     /strain="Cl - Brenner"
                     /db_xref="taxon:5693"
                     /clone="n1059.r"
                     /cell_type="epimastigote"
                     /clone_lib="T. Cruzi epimastigote normalised cDNA Library"
                     /note="Site_1: EcoRI; Site_2: NotI; cDNA library
                     constructed with oligo dt primed  epimastigote mRNA and
                     cloned in pt7t318D  phagemid with modified polylinker"
ORIGIN

  Query Match             8.5%;  Score 38.8;  DB 9;  Length 775;
  Best Local Similarity   57.4%;  Pred. No. 0.84;
  Matches   70;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        245 TCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATG 304
              |||    ||||| |||||  |  ||||| |     ||| ||   ||  |||||  |  |
Db         97 TCGCAGTGGCAGCGAATTCGAGATTTGCATTCTGGGAGAAACACGGTCATGCCGTCTGTT 156

Qy        305 CCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAA 364
              || ||| |||  | |  |    ||||   |   |||  |||| ||  ||| |  |||||
Db        157 CCTGTAGATCCAACACTGTCGTCGGTGCGACCTCAGTCATTACACACGAAGCTGGTGGAA 216

Qy        365 AT 366
              ||
Db        217 AT 218


RESULT 6
```

```
AZ578608/c
LOCUS       AZ578608                288 bp    DNA     linear   GSS 08-DEC-2000
DEFINITION  25d10 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
            sp. NGR234 genomic clone 25d10, genomic survey sequence.
ACCESSION   AZ578608
VERSION     AZ578608.1  GI:11606060
KEYWORDS    GSS.
SOURCE      Rhizobium sp. NGR234
  ORGANISM  Rhizobium sp. NGR234
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
REFERENCE   1  (bases 1 to 288)
  AUTHORS   Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
  TITLE     Genetic snapshots of the Rhizobium species NGR234 genome
  JOURNAL   Genome Biol. 1 (6), RESEARCH0014 (2000)
  MEDLINE   21114532
   PUBMED   11178268
COMMENT     Contact: Virginie Viprey
            Laboratoire de Biologie Moleculaire des Plantes Superieures
            University of Geneva
            1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
            Tel: +44(0)1603450000
            Fax: +44(0)1603450045
            Email: virginie.viprey@bbsrc.ac.uk
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism="Rhizobium sp. NGR234"
                     /mol_type="genomic DNA"
                     /strain="ANU265"
                     /db_xref="taxon:394"
                     /clone="25d10"
                     /clone_lib="Shot-gun genomic library of Rhizobium strain
                     ANU265"
                     /note="Vector: M13; derivative strain of NGR234 cured of
                     pNGR234a"
ORIGIN

  Query Match             8.3%;  Score 38;  DB 28;  Length 288;
  Best Local Similarity   50.0%;  Pred. No. 1.1;
  Matches   95;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        204 GTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTA 263
              ||||||   |   || |||||        |||    |   |   | |||   || ||  |
Db        238 GTCCGTGTCTGTTCAGGAAGGCAAGGGCAATATCTCCGGAACGGCGCAGTTCGGAAAGAA 179

Qy        264 TAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGTATATCGCAAAGCGC 323
                ||   |||   |    |||||  ||||||   |||||  | |  | |  |  ||   ||
Db        178 CCACAAGGCGATCACCCTGCAAAAGGGCAAGTACAACGCTTCGGGCACGTTCCAGTTCGG 119

Qy        324 TTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGGAAATAAGGCCAATATTACCCA 383
              ||||||  || |   |    |    ||||  | ||    ||||||    |||  |     ||
Db        118 TTACGGCAACAAATCGCTGACCCACCAGGACGGAGAAGGAAATCTCTCCATCACAGTGCA 59

Qy        384 GTACGGTACG 393
              | ||  |  ||
Db         58 GGACTGATCG 49


RESULT 7
BP007561
LOCUS       BP007561                675 bp    mRNA    linear   EST 15-MAR-2002
DEFINITION  BP007561 Nori Satoh unpublished cDNA library, young adult Ciona
            intestinalis cDNA clone ciad45a18 5', mRNA sequence.
ACCESSION   BP007561
VERSION     BP007561.1  GI:19499038
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .675
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="ciad45a18"
                     /tissue_type="whole animal"
                     /dev_stage="young adult"
                     /clone_lib="Nori Satoh unpublished cDNA library, young
                     adult"
ORIGIN

  Query Match             8.3%;  Score 37.8;  DB 12;  Length 675;
  Best Local Similarity   50.3%;  Pred. No. 1.6;
  Matches   93;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        229 AATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACG 288
              ||||  |       ||   ||    | |||  | | | | || | |  | |  ||| |||
Db        257 AATACTTTCAATGAAAAAGAAATGTCAGAAACTAAAACCTGTACTGAAACAGAGAAAAAA 316

Qy        289 GGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATC 348
               ||    ||   |  | || | |||   | |      |  | |||   ||  ||   |
Db        317 TGCTCCACCTCTGCAGGCAATGTATTAGATACAAAAGAAAGCAATTTAGCTTCTGAAACT 376

Qy        349 CAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTT 408
              || | |  ||||| |||||| | |||| |    ||| | | ||   | |  ||| ||| |
Db        377 CAAATAATTTCTGCAAATAATGACAATTTCCAGCAGAATGATAACGACACTACACCAGAT 436

Qy        409 GTAGT 413
              | | |
Db        437 GAACT 441


RESULT 8
AV896333/c
LOCUS       AV896333                715 bp    mRNA    linear   EST 09-NOV-2001
DEFINITION  AV896333 Nori Satoh unpublished cDNA library, young adult Ciona
            intestinalis cDNA clone rciad45a18 3', mRNA sequence.
ACCESSION   AV896333
VERSION     AV896333.1  GI:16885431
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="rciad45a18"
                     /tissue_type="whole animal"
                     /dev_stage="young adult"
```

/clone_lib="Nori Satoh unpublished cDNA library, young adult"
ORIGIN

Query Match 8.3%; Score 37.8; DB 9; Length 715;
Best Local Similarity 50.3%; Pred. No. 1.7;
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```
Qy        229 AATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACG 288
              |||| |       ||   ||   | ||| | | | | || | |  | |  ||| |||
Db        620 AATACTTTCAATGAAAAAGAAATGTCAGAAACTAAAACCTGTACTGAAACAGAGAAAAAA 561

Qy        289 GGCAATGCCAACGATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATC 348
               ||   ||   |  | || | |||   | |       |  | |||   ||  ||   |
Db        560 TGCTCCACCTCTGCAGGCAATGTATTAGATACAAAAGAAAGCAATTTAGCTTCTGAAACT 501

Qy        349 CAGAAAGGTTCTGGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTT 408
              || | |  ||||| |||||| | |||| |     ||| | | ||   | |  ||| ||| |
Db        500 CAAATAATTTCTGCAAATAATGACAATTTCCAGCAGAATGATAACGACACTACACCAGAT 441

Qy        409 GTAGT 413
              | | |
Db        440 GAACT 436
```

RESULT 9
BU213938
LOCUS BU213938 601 bp mRNA linear EST 25-NOV-2002
DEFINITION 603108484F1 CSEQCHN04 Gallus gallus cDNA clone ChEST51g22 5', mRNA sequence.
ACCESSION BU213938
VERSION BU213938.1 GI:25389321
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES Location/Qualifiers
source 1. .601
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST51g22"
/tissue_type="whole embryo"
/dev_stage="20-21"
/lab_host="DH10B"
/clone_lib="CSEQCHN04"
/note="Organ: whole embryo; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."
ORIGIN

Query Match 8.2%; Score 37.6; DB 13; Length 601;
Best Local Similarity 50.6%; Pred. No. 1.8;
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy         87 AGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTAT 146
              || | | ||| |    ||||||| |  |  | || ||  |  | | |  | | |  | ||
Db        172 AGCAAAAAATCTACGCGTAAATGCAGGAGACCAGCCTGGAGCTGACCTAGGGCCTCTAAT 231

Qy        147 TGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGTC 206
               |||      | || |||  |     |  |     |    ||| |||||  | ||  |    |
Db        232 CAGTCCCCAAGCCAAGGAAAGGGTTTGCCACCTGATTCAGAAAGGAGTAGAAGAAGGTGC 291

Qy        207 CGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTATAA 266
              |  | ||   |  || ||| | |||| |   | ||| | | |  ||  || |||| | |
Db        292 CAGTCTTCTTCTGGATGGACGTAATATTAAAGTGAAGGGCTATGAGAACGGCAATTCTGA 351
```

RESULT 10
CC310289/c
LOCUS CC310289 1214 bp DNA linear GSS 14-MAY-2003
DEFINITION TAM32-32B18_EC1.1 TAM32 Gallus gallus genomic clone TAM32-32B18, genomic survey sequence.
ACCESSION CC310289
VERSION CC310289.1 GI:30703437
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1214)
AUTHORS Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 152000 Std Error: 0.00
Seq primer: EC1 TACGACTCACTATAGGGCG
Class: BAC ends
High quality sequence start: 20
High quality sequence stop: 781.
FEATURES Location/Qualifiers
source 1. .1214
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="TAM32-32B18"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="TAM32"
/note="Vector: pECBAC1; Site_1: EcoRI; Site_2: EcoRI; TAM32 Female Chicken library - for library and clone ordering information: http://www.hbz.tamu.edu"
ORIGIN

Query Match 8.2%; Score 37.4; DB 28; Length 1214;
Best Local Similarity 47.3%; Pred. No. 2.6;
Matches 113; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy         71 ATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTA 130
              ||  | ||| | ||||| | || | |||   ||  | | |   | | |||| ||||   |
Db       1027 ATATTTTGGATGGTTCAAAGTAAATTTTGAGGGGCATTTACAGATGGAAGTTTTCAAAAA 968
```

```
Qy        131 ATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAG 190
Db        967 ATCCAGGGAACTTTATTAGTGGAGTAGAGAACCATCCTGGTTAATTCCTGGTCCTTGAGT 908

Qy        191 GATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACC 250
Db        907 TAGCAAAATGGGATACCTTAAGAGAACAAAAGAAAACCTGGCCATTTGCGGTCTCCCAAA 848

Qy        251 AGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGATGCCAGT 309
Db        847 TACCAGGAAACAAACTATTTATCTATCCTGCTCAAACAGTTCAAGGAAGCTAATCTAAT 789
```

RESULT 11
BU387712
LOCUS BU387712 708 bp mRNA linear EST 28-NOV-2002
DEFINITION 603857374F1 CSEQCHN75 Gallus gallus cDNA clone ChEST863c20 5', mRNA sequence.
ACCESSION BU387712
VERSION BU387712.1 GI:25895713
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 708)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES Location/Qualifiers
source 1. .708
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST863c20"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHN75"
/note="Organ: trunks; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: Not1; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."
ORIGIN

Query Match 8.1%; Score 36.8; DB 13; Length 708;
Best Local Similarity 50.6%; Pred. No. 3.4;
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy         89 AATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTATTG 148
Db        17 AAGAGAATCTACGCGTAAATGCAGGAGACCAGCCTGGAGCTGACCTAGGGCCTCTAATCA 76

Qy        149 GTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAACTATTGTCCG 208
Db         77 GTCCCCAAGCCAAGGAAAGGGTTTGCCACCTGATTCAGAAAGGAGTAGAAGAAGGTGCCA 136

Qy        209 TTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGACCAGGCAGGGAATTAT 264
Db        137 GTCTTCTTCTGGATGGACGTAATATTAAAGTGAAGGGCTATGAGAACGGCAATTTT 192
```

RESULT 12
AZ529672
LOCUS AZ529672 924 bp DNA linear GSS 03-NOV-2000
DEFINITION ENTBM10TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
ACCESSION AZ529672
VERSION AZ529672.1 GI:11082756
KEYWORDS GSS.
SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 924)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 33
High quality sequence stop: 745.
FEATURES Location/Qualifiers
source 1. .924
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
ORIGIN

Query Match 8.1%; Score 36.8; DB 28; Length 924;
Best Local Similarity 50.6%; Pred. No. 3.7;
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy         73 GATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAAT 132
Db        740 GATGTTACACTTAGAAGATTTATATAAGGGGTAAAACTATTAGAGGGTATTACATTAGAA 799

Qy        133 CAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGA 192
```

```
Db        800 AAAGACACATTATTTGTCATCATTGTCATGGGTACTGGGAAGGAGATACACCACGTATGA 859

Qy        193 TCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAAAGTCGA 248
              |  |||  | |||||  ||| |||   ||||  || | | |    |||| |   ||
Db        860 TGTAAAGAAATGTCCAGTATGTCATTGGAAGATGGGAGTTAAAACGCGAGACAAGA 915


RESULT 13
CC211905/c
LOCUS       CC211905               1550 bp    DNA     linear   GSS 12-MAY-2003
DEFINITION  CH261-18606_Sp6.1 CH261 Gallus gallus genomic clone CH261-18606,
            genomic survey sequence.
ACCESSION   CC211905
VERSION     CC211905.1  GI:30530573
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1550)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 339
            High quality sequence stop: 1106.
FEATURES             Location/Qualifiers
     source          1. .1550
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-18606"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 28;  Length 1550;
  Best Local Similarity   59.8%;  Pred. No. 5.9;
  Matches   61;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         97 TTTGCGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTC 156
              ||||  |   | ||||||  |||  ||  ||| | ||||  ||  ||    |||| ||   ||
Db       1185 TTTGGGTGGACTGAATTGAGCCTGTGTTCTTGTAATGTGGAAGCACAAATTGTTCCTCTC 1126

Qy        157 GGCACGGATAATAGTGCCAGAGTACGCCAGGAAGGATCAAAA 198
               || | |  ||||    |   |||  |    |||||||  ||||
Db       1125 AGCTCAGGAAATATCTCTGTAGTTCAGGGGGAAGGAAAAAAA 1084


RESULT 14
BH528896/c
LOCUS       BH528896               499 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGPO61TR BOGP Brassica oleracea genomic clone BOGPO61, genomic
            survey sequence.
ACCESSION   BH528896
VERSION     BH528896.1  GI:17744694
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGPO61TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGPO61"
                     /clone_lib="BOGP"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             7.9%;  Score 36;  DB 28;  Length 499;
  Best Local Similarity   46.4%;  Pred. No. 5.4;
  Matches  117;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        183 CCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCGAA 242
              || | || || ||||   |||  | || |||  | | | ||  || || |         |
Db        470 CCGGAAACGAGCAAATGCATTCACGGTGATTGAAGATGCAGAGGGCAACATGGTCTATCA 411

Qy        243 AGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAACGA 302
              ||  || |||  |||||      | | ||| |||| || |  |       |||  | |  ||
Db        410 AGAAGATCAGATAGGGAGAGTGATCATTGAGTATTTTCATGAGCTATTCAAGACTATTGA 351

Qy        303 TGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCTGG 362
              | ||  | |  | ||| ||      ||  |    || | ||  |      ||   |
Db        350 GGGCAACAGAGAGAAAACCGTAATGCGTGCTCTATCACCAATGGTGAGCGCAGAAACGAA 291

Qy        363 AAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAACA 422
              | |||   | ||    ||   | | | | |||  | ||  |   || |    |  ||
Db        290 CGAGAAGCTAATTACAGTCCCAAAAGCTGCAGAGATCAAAGAGGCCCTATTCTCTATTCA 231

Qy        423 GTCGCATATGGC 434
              |  | |  ||
Db        230 TGCCGACAAAGC 219


RESULT 15
BG226083
LOCUS       BG226083               579 bp    mRNA    linear   EST 09-MAY-2001
DEFINITION  kq08h12.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to
            WP:ZK686.3 CE00457 ;, mRNA sequence.
ACCESSION   BG226083
VERSION     BG226083.1  GI:12713638
KEYWORDS    EST.
SOURCE      Strongyloides stercoralis
  ORGANISM  Strongyloides stercoralis
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Panagrolaimoidea; Strongyloididae; Strongyloides.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
```

```
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Thomas Nutman and colleagues of
            NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
            University Genome Sequencing Center St. Louis.
            High quality sequence stop: 417.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Strongyloides stercoralis"
                     /mol_type="mRNA"
                     /strain="Rhabditiform larvae obtained from gerbils"
                     /db_xref="taxon:6248"
                     /lab_host="XL-1 Blue MRF' (Stratagene)"
                     /clone_lib="TBN95TM-SSR"
                     /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
                     rhabditiform larvae which had been isolated from gerbils
                     experimentally infected with larvae originally isolated
                     from experimentally infected dogs. cDNA was constructed
                     and, using adaptors, was cloned unidirectionally into the
                     vector from the EcoRI site to the XhoI site. The library
                     has an unamplified titer of 1 x 10E5 pfu/ml and an
                     amplified, undiluted titer of 9 x 10E11 pfu/ml. The
                     average insert size of the unamplifed library is 675 bp
                     (range, 100-1700)."
ORIGIN

  Query Match             7.9%;  Score 36;  DB 12;  Length 579;
  Best Local Similarity   49.0%;  Pred. No. 5.6;
  Matches   96;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          7 AACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACCGCGACA 66
              || ||| | |  ||| | ||| |          | ||| || | ||||    |  | | |
Db        271 AAAAAACTTTATTTTGTTATGGTTGATTATGCAGATGCTCCAGAGATTTTTGCTTCAATA 330

Qy         67 AATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAGTCTTCA 126
              |||  | ||  ||||  ||       |||||| | |  || | ||  || || |||
Db        331 AATCTTAATACAGCTCCCGCATTTATACATTTTCCAGCTAAAGGATCTAGGAAATCTGAT 390

Qy        127 TTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTACGCCAG 186
                ||    ||  |     |||  |  | | |   | |||  |   || ||| |     |
Db        391 GATACAATGGATGTTCAAATTAATGGATTTGAAGCTGATGTTTTGGCAAGATTTGTTAAA 450

Qy        187 GAAGGATCAAAACTAT 202
              ||| || |  || |||
Db        451 GAAAGAACTGAAATAT 466


Search completed: March 18, 2004, 20:11:11
Job time : 2038.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		456	100.0	456	6	AX814811	AX814811 Sequence
2		456	100.0	648	1	ECOCSGAA	L04979 Escherichia
3		454.4	99.6	4680	1	ECCSGABDG	X90754 E.coli csgG
4		454.4	99.6	10346	1	AE000205	AE000205 Escherich
5		454.4	99.6	15047	1	D90741	D90741 Escherichia
6		395	86.6	1711	1	AF275733	AF275733 Escherich
7		395	86.6	10190	1	AE005315	AE005315 Escherich
8		395	86.6	327773	1	AP002554	AP002554 Escherich
9		382.2	83.8	306358	1	AE016759	AE016759 Escherich
10		340.8	74.7	10370	1	AE015131	AE015131 Shigella
11		340.8	74.7	292504	1	AE016981	AE016981 Shigella
12	c	311.4	68.3	437	1	AF237726	AF237726 Shigella
13		259.2	56.8	5103	1	STAJ2301	AJ002301 Salmonell
14		259.2	56.8	22411	1	AE008749	AE008749 Salmonell
15		257.6	56.5	254050	1	AL627269	AL627269 Salmonell
16	c	257.6	56.5	301983	1	AE016840	AE016840 Salmonell
17		254.4	55.8	456	6	I44909	I44909 Sequence 58
18		254.4	55.8	2067	1	SEU43280	U43280 Salmonella
19		253.6	55.6	1048	1	STAGFBA	AJ000514 Salmonell
20		252.8	55.4	2883	1	ESA515702	AJ515702 Enterobac
21		246.4	54.0	2889	1	CSP515700	AJ515700 Citrobact
22		230.8	50.6	2920	1	CFR515701	AJ515701 Citrobact
23		186	40.8	19201	1	D90742	D90742 Escherichia
24		173.4	38.0	361	6	I44908	I44908 Sequence 56
25		121.8	26.7	230	1	SEU53207	U53207 Salmonella
26		83.8	18.4	1212	1	ECO131756	AJ131756 Escherich
27		78	17.1	78	6	AX814809	AX814809 Sequence
28		72	15.8	72	6	AX814798	AX814798 Sequence
29		66	14.5	66	6	AX814808	AX814808 Sequence
30		43.4	9.5	301214	1	AE016786	AE016786 Pseudomon
31		43.2	9.5	85600	9	AC138647	AC138647 Homo sapi
32	c	42.6	9.3	135259	9	HS127B20	Z83838 Human DNA s
33	c	42.6	9.3	165074	9	AC093519	AC093519 Homo sapi
34		41.6	9.1	181047	2	AC016765	AC016765 Homo sapi
35		41.6	9.1	198599	9	AC108448	AC108448 Homo sapi
36		41.2	9.0	159070	2	AC128114	AC128114 Rattus no
37	c	41.2	9.0	186562	2	AL451079	AL451079 Homo sapi
38	c	41.2	9.0	200368	9	AC096541	AC096541 Homo sapi
39		41.2	9.0	241394	2	AC132789	AC132789 Rattus no
40		40.8	8.9	165074	9	AC093519	AC093519 Homo sapi
41	c	40.6	8.9	236533	2	AC103433	AC103433 Rattus no
42		40.2	8.8	183458	2	AC129554	AC129554 Mus muscu
43	c	40	8.8	180995	9	AC117500	AC117500 Homo sapi
44		40	8.8	187615	2	AC020724	AC020724 Homo sapi
45		39.8	8.7	999	1	AY077581	AY077581 Bradyrhiz

ALIGNMENTS

RESULT 1
AX814811
LOCUS AX814811 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Bjoerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
TITLE Peptides
JOURNAL Patent: WO 03064446-A 15 07-AUG-2003;

Hansa Medical Research Aktiebolag (SE)

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .456
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE83541.1"
                     /db_xref="GI:39104002"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
ORIGIN
```

Query Match 100.0%; Score 456; DB 6; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.6e-122;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

```
RESULT 2
ECOCSGAA
LOCUS       ECOCSGAA                  648 bp    DNA     linear   BCT 13-JUL-1993
DEFINITION  Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION   L04979
VERSION     L04979.1  GI:290424
KEYWORDS    csgA gene; curli organelle; fibronectin-binding protein.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Olsen,A., Arnqvist,A., Hammar,M., Sukupolvi,S. and Normark,S.
  TITLE     The RpoS sigma factor relieves H-NS-mediated transcriptional
            repression of csgA, the subunit gene of fibronectin-binding curli
            in Escherichia coli
  JOURNAL   Mol. Microbiol. 7 (4), 523-536 (1993)
  MEDLINE   93211294
   PUBMED   8459772
```

```
COMMENT     On Jun 11, 1993 this sequence version replaced gi:145630.
            Original source text: Escherichia coli (sub_strain W3110, strain
            K-12) (library: Kohara) DNA.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
                     /tissue_lib="Kohara"
     gene            83. .538
                     /gene="csgA"
     CDS             83. .538
                     /gene="csgA"
                     /note="precursor protein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="curlin subunit"
                     /protein_id="AAA23616.1"
                     /db_xref="GI:290425"
                     /translation="MKLLKVEAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     83. .142
                     /gene="csgA"
                     /evidence=experimental
     mat_peptide     143. .535
                     /gene="csgA"
                     /product="curlin subunit"
                     /evidence=experimental
ORIGIN
```

Query Match 100.0%; Score 456; DB 1; Length 648;
Best Local Similarity 100.0%; Pred. No. 9e-122;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

```
RESULT 3
ECCSGABDG
```

```
LOCUS       ECCSGABDG               4680 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION   X90754
VERSION     X90754.1  GI:1147558
KEYWORDS    csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
            orfC gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
  TITLE     Expression of two csg operons is required for production of
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match             99.6%;  Score 454.4;  DB 1;  Length 4680;
  Best Local Similarity   99.8%;  Pred. No. 3.5e-121;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       3729 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 3788

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3789 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 3848

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3849 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 3908

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3909 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 3968

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3969 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 4028

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4029 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 4088

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4089 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 4148

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       4149 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 4184


RESULT 4
AE000205
LOCUS       AE000205               10346 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
            genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1  GI:1787265
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10346)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10346)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10346)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   complement(<1. .48)
                     /note="IS3"
     promoter        <1. .5
                     /note="factor Sigma70; predicted +1 start at 1094689"
     promoter        5. .32
                     /note="factor Sigma70; predicted +1 start at 1094716"
     gene            69. .392
                     /gene="b1028"
     CDS             69. .392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS
                     LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD
                     AKVFK"
     promoter        275. .299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
     gene            389. .1375
```

```
                     /gene="ycdU"
                     /note="synonym: b1029"
     CDS             389. .1375
                     /gene="ycdU"
                     /function="orf; Unknown"
                     /note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCGE
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELFLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
     protein_bind    1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
     promoter        1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
     gene            1494. .1745
                     /gene="b1030"
     CDS             1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
     gene            1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
     CDS             1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
     gene            complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
     tRNA            complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
     promoter        2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
     promoter        2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
     promoter        2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
     gene            2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
     CDS             2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     promoter        3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
     promoter        3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
     promoter        3396. .3423

  Query Match             99.6%;  Score 454.4;  DB 1;  Length 10346;
  Best Local Similarity   99.8%;  Pred. No. 3.9e-121;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       8993 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 9052

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9053 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 9112

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9113 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 9172

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9173 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 9232

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9233 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 9292

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 9352

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9353 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 9412

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       9413 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 9448


RESULT 5
D90741
LOCUS       D90741                 15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
```

KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdE; cls; nov; mdoG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K., Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL DNA Res. 3 (3), 137-155 (1996)
MEDLINE 97061202
PUBMED 8905232
REFERENCE 2
AUTHORS Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 15047)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
COMMENT Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.

FEATURES Location/Qualifiers

source 1. .15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the initiation site of ThrA (0 min.).~This clone is from Kohara lambda miniset library."

CDS <1. .276
/note="unnamed protein product; ORF_ID:o230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"

CDS 1152. .2090
/note="ORF_ID:o231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in proB-proA intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"

CDS 2145. .2882
/note="unnamed protein product; ORF_ID:o231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER ILNVSPRRLLNFLESRGMAPIAEFADL"

CDS 2906. .3460
/note="ORF_ID:o231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=11
/product="Hypothetical protein HI1543"
/protein_id="BAA35825.1"
/db_xref="GI:4062608"
/translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW RTMAPLTRDAISAMWDELEEDSEE"

CDS 3514. .4053
/note="unnamed protein product; ORF_ID:o231#4"
/codon_start=1
/transl_table=11
/protein_id="BAA35826.1"
/db_xref="GI:4062609"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW LAARSAKTAHREQEIKNKA"

gene complement(4117. .4950)
/gene="csgG"

CDS complement(4117. .4950)
/gene="csgG"
/note="ORF_ID:o231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component CsgG precursor."
/protein_id="BAA35827.1"
/db_xref="GI:4062610"
/translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE

```
                S"
     gene       complement(4977. .5393)
                /gene="csgF"
     CDS        complement(4977. .5393)
                /gene="csgF"
                /note="ORF_ID:o231#6
                similar to SwissProt Accession Number P52104"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgF precursor."
                /protein_id="BAA35828.1"
                /db_xref="GI:4062611"
                /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene       complement(5418. .5807)
                /gene="csgE"
     CDS        complement(5418. .5807)
                /gene="csgE"
                /note="ORF_ID:o231#7
                similar to SwissProt Accession Number P52105"
                /codon_start=1
                /transl_table=11
                /product="Curli production assembly/transport component
                CsgE precursor."
                /protein_id="BAA35829.1"
                /db_xref="GI:4062612"
                /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                IQTEEALNRRQINQALLSTGDLAHDEF"
     gene       complement(5812. .6462)
                /gene="csgD"
     CDS        complement(5812. .6462)
                /gene="csgD"
                /note="ORF_ID:o231#8
                similar to SwissProt Accession Number P52106"
                /codon_start=1
                /transl_table=11
                /product="Probable csgAB operon transcriptional regulatory
                protein."
                /protein_id="BAA35830.1"
                /db_xref="GI:4062613"
                /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene       7217. .7672
                /gene="csgB"
     CDS        7217. .7672

  Query Match          99.6%;  Score 454.4;  DB 1;  Length 15047;
  Best Local Similarity  99.8%;  Pred. No. 4.1e-121;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       7713 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 7772

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7773 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 7832

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7833 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 7892

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7893 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 7952

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       7953 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 8012

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8013 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 8072

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8073 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 8132

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       8133 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 8168


RESULT 6
AF275733
LOCUS       AF275733               1711 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
            CsgA protein (csgA) genes, complete cds.
ACCESSION   AF275733
VERSION     AF275733.1  GI:14039399
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Mutations in the csgD promoter associated with variations in curli
            expression in certain strains of Escherichia coli O157:H7
  JOURNAL   Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
  MEDLINE   21218556
   PUBMED   11319125
REFERENCE   2  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
            Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES             Location/Qualifiers
     source          1. .1711
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="43895 Red Variant"
                     /serotype="O157:H7"
                     /db_xref="taxon:562"
     gene            757. .1212
                     /gene="csgB"
     CDS             757. .1212
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="AAK53211.1"
                     /db_xref="GI:14039400"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1253. .1711
                     /gene="csgA"
     CDS             1253. .1711
                     /gene="csgA"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="AAK53212.1"
                     /db_xref="GI:14039401"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match          86.6%;  Score 395;  DB 1;  Length 1711;
```

```
  Best Local Similarity   92.8%;  Pred. No. 6.3e-104;
  Matches  426;  Conservative    0;  Mismatches   30;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1253 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 1312

Qy         61 GGTGTTGTTCCTCAGTACGGCGGC---GGCGGTAACCACGGTGGTGGCGGTAATAATAGC 117
              |||||||||||||||||||||||   |||||||||||||||||||||||||||||||||| |||
Db       1313 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 1372

Qy        118 GGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTG 177
              ||||| ||||| |||||||| ||||| |||||||||| ||||||||||||||||||||||
Db       1373 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 1432

Qy        178 CAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCA 237
              ||| ||||||| |||||||||| | ||||||||||||||||| || ||||| ||||||
Db       1433 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 1492

Qy        238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAAC 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1493 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 1552

Qy        298 AGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGT 357
              ||||| |||||||||||||||||||| ||| | ||| | ||||| |||||||| ||||||
Db       1553 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 1612

Qy        358 GGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAG 417
              || |||||||||| || |||||||||||||||||||| ||| |||||||||| ||||||
Db       1613 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 1672

Qy        418 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||
Db       1673 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 1711
```

```
RESULT 7
AE005315
LOCUS       AE005315               10190 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
            of 155.
ACCESSION   AE005315 AE005174
VERSION     AE005315.1  GI:12514572
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10190
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            617. .1072
                     /gene="csgB"
     CDS             617. .1072
                     /gene="csgB"
                     /function="structural component; Cell exterior
                     constituents: Surface structures"
                     /note="Residues 1 to 151 of 151 are 100.00 pct identical
                     to residues 1 to 151 of 151 from Escherichia coli K-12
                     Strain MG1655: B1041"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAG55787.1"
                     /db_xref="GI:12514573"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1113. .1571
                     /gene="csgA"
     CDS             1113. .1571
                     /gene="csgA"
                     /function="regulator; Cell exterior constituents: Surface
                     structures"
                     /note="Residues 1 to 152 of 152 are 96.71 pct identical to
                     residues 1 to 151 of 151 from Escherichia coli K-12 Strain
                     MG1655: B1042"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin major subunit, coiled surface structures;
                     cryptic"
                     /protein_id="AAG55788.1"
                     /db_xref="GI:12514574"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
     gene            1630. .1962
                     /gene="csgC"
     CDS             1630. .1962
                     /gene="csgC"
                     /function="putative structure; Not classified"
                     /note="Residues 1 to 110 of 110 are 98.18 pct identical to
                     residues 1 to 110 of 110 from Escherichia coli K-12 Strain
                     MG1655: B1043"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="AAG55789.1"
                     /db_xref="GI:12514575"
                     /translation="MNALLLLAALSSQITFNTTQQGDMYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     gene            2083. .2394
                     /gene="Z1678"
     CDS             2083. .2394
                     /gene="Z1678"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 103 of 103 are 98.05 pct identical to
                     residues 1 to 103 of 103 from Escherichia coli K-12 Strain
                     MG1655: B1044"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55790.1"
                     /db_xref="GI:12514576"
                     /translation="MFRPFLNSLMLGSLFFPFIAIAGSTAQGGVIHFYGQIVEPACDV
                     STQSSPVEMNCPQNGSVPGKTYSSKALMSGNVKNAQIASVKVQYLDKQKKLAVMNIEY
```

```
                N"
gene            2489. .3022
                /gene="Z1679"
CDS             2489. .3022
                /gene="Z1679"
                /function="orf; Unknown function"
                /note="Residues 1 to 177 of 177 are 100.00 pct identical
                to residues 1 to 177 of 177 from Escherichia coli K-12
                Strain MG1655: B1045"
                /codon_start=1
                /transl_table=11
                /product="putative polyprotein"
                /protein_id="AAG55791.1"
                /db_xref="GI:12514577"
                /translation="MKTRIHVVQGDITKLAVDVIVNAANPSLMGGGGVDGAIHRAAGP
                ALLDACLKVRQQQGDCPTGHAVITLAGDLPAKAVVHTVGPVWRGGEQNEDQLLQDAYL
                NSLRLVAANSYTSVAFPAISTGVYGYPRAAAAEIAVKTVSEFITRHALPEQVYFVCYD
                EENAHLYERLLTQQGDE"
gene            2964. .4445
                /gene="ymdC"
                /note="synonym: Z1680"
CDS             2964. .4445
                /gene="ymdC"
                /function="putative enzyme; Not classified"
                /note="Residues 1 to 493 of 493 are 99.59 pct identical to
                residues 1 to 493 of 493 from Escherichia coli K-12 Strain
                MG1655: B1046"
                /codon_start=1
                /transl_table=11
                /product="putative synthase"
                /protein_id="AAG55792.1"
                /db_xref="GI:12514578"
                /translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPGQCGLFPL
                EKSLDAPAARYRLAEMSEHTLDVQYYIWQDDMSGRLLFSALLAAAKRGVRVRLLLDDN
                NTPGLDDILRLLDSHPRIEVRLFNPFSFRLLRPLGYITDFSRLNRRMHNKSFTVDGVV
                TLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWYCKSVSPLQQVLDV
                PEGEMADRIELPASWHNDAMTHRYLRKMESSPFINHLVDGTLPLIWAKTRLLSDDPAK
                GEGKAKRHSLLPQRLFDIMGSPSERIDIISSYFVPTRAGVAQLLRMVRKXVKIAILTN
                SLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREQSSTLHDRGITGNSGASLHAKT
                FSIDGKTVFIGSFNFDPRSTLLNTEMGFVIESETLAQLIDKRFIQSQYDAAWQLRLDR
                WGRINWVDRHAKKEIVLKKEPATSFWKRVMVRLASILPVEWLL"
gene            complement(4453. .5610)
                /gene="ymdD"
                /note="synonym: Z1681"
CDS             complement(4453. .5610)
                /gene="ymdD"
                /function="orf; Unknown function"
                /note="Residues 1 to 385 of 385 are 99.22 pct identical to
                residues 1 to 385 of 385 from Escherichia coli K-12 Strain
                MG1655: B1047"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG55793.1"
                /db_xref="GI:12514579"
                /translation="MNPVPAQREYFXDSIRAWLMLLGIPFHISLIYSSHTWHVNSAEP
                SLWLTLFNDFIHSPRMQVPFVISGYPSYMLFLRYPLKKWWKVRVERVGIPMLTAIPLL
                TLPQFIMLQYVKGKAESWPGLSLYDKYNTLAWELISHLWFLLVLVVMTTLCVWIFKRI
                RNNLENSDKTNKKFSMVKLSVIFLCLGIGYAVIRRTIFIVYPPILSNGTFNFIVMQTL
                FYLPFFILGALAFIFPPHLKALFTTPSRGCTLAAALAFVAYLLNQRYGSGDAWMYETES
                VITMVLGLWMVNVVFSFGHRLLNFQSARVTYFVNASLFIYLVHHPLTLFFGAYITPHI
                TSNWLGFLCGLIFVVGIAIILYEIHLRIPLLKFLFSGKPVVKRENDKAPAR"
gene            6003. .7538
                /gene="mdoG"
                /note="synonym: Z1683"
CDS             6003. .7538
                /gene="mdoG"
                /function="enzyme; Osmotic adaptation"
                /note="Residues 1 to 511 of 511 are 99.80 pct identical to
                residues 1 to 511 of 511 from Escherichia coli K-12 Strain
                MG1655: B1048"
                /codon_start=1
                /transl_table=11
                /product="periplasmic glucans biosynthesis protein"
                /protein_id="AAG55794.1"
                /db_xref="GI:12514580"
                /translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAQSLAGKGYEAPK
                SNLPSVFRDMKYADYQQIQFNHDKAYWNNLKTPFKLEFYHQGMYFDTPVKINEVTATA
                VKRIKYSPDYFTFGDVQHDKDTVKDLGFAGFKVLYPINSKDKNDEIVSMLGASYFRVI
                GAGQVYGLSARGLAIDTALPSGEEFPRFKEFWIERPKPTDKRLTIYALLDSPRATGAY
                KFVVMPGRDTVVDVQSKIYLRDKVGKLGVAPLTSMFLFGPNQPSPANNYRPELHDSNG
                LSIHAGNGEWIWRPLNNPKHLAVSSFSMENPQGFGLLQRGRDFSRFEDLDDRYDLRPS
                AWVTPKGEWGKGSVELVEIPTNDETNDNIVAYWTPDQLPEPGKEMNFKYTITFSRDED
                KLHAPDNAWVQQTRRSTGDVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDTPVTAQTS
                IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDAKKTTEMRAALVNADQTLSETWSYQLP
                ANE"
gene            7531. .10074
                /gene="mdoH"
                /note="synonym: Z1684"
CDS             7531. .10074
                /gene="mdoH"
                /function="enzyme; Osmotic adaptation"
```

Query Match 86.6%; Score 395; DB 1; Length 10190;
Best Local Similarity 92.8%; Pred. No. 8.1e-104;
Matches 426; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       1113 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 1172

Qy         61 GGTGTTGTTCCTCAGTACGGCGGC---GGCGGTAACCACGGTGGTGGCGGTAATAATAGC 117
              |||||||||||||||||||||||   |||||||||||||||||||||||||||||| |||
Db       1173 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 1232

Qy        118 GGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTG 177
              ||||| ||||| |||||||| ||||| |||||||||| ||||||||||||||||||||||
Db       1233 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 1292

Qy        178 CAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCA 237
              ||| ||||||| |||||||||||   | ||||||||||||||||| || ||||| ||||||
Db       1293 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 1352

Qy        238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAAC 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1353 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 1412

Qy        298 AGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGT 357
              ||||| |||||||||||||||||| ||| | ||| | ||||| |||||||| ||||||
Db       1413 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 1472

Qy        358 GGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAG 417
              || |||||||||| || |||||||||||||||||||| ||| |||||||| ||||||
Db       1473 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 1532

Qy        418 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||
Db       1533 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 1571
```

```
RESULT 8
AP002554
LOCUS       AP002554              327773 bp    DNA     linear   BCT 21-DEC-2002
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION   AP002554 BA000007
VERSION     AP002554.1  GI:13360491
KEYWORDS    .
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
```

```
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
   PUBMED   10734605
REFERENCE   2
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
   PUBMED   11108008
REFERENCE   3
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
   PUBMED   11111050
REFERENCE   4
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
   PUBMED   11258796
REFERENCE   5  (bases 1 to 327773)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .327773
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            150. .2258
                     /gene="ECs1032"
     CDS             150. .2258
                     /gene="ECs1032"
                     /note="similar to YCBY_ECOLI gi|1787181 percent identity
                     99 in 702 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative oxidoreductase"
                     /protein_id="BAB34455.1"
                     /db_xref="GI:13360492"
                     /translation="MNSLFASTARGLEELLKTELENLGAVECQVVQGGVHFKGDTRLV
                     YQSLMWSRLASRIMLPLGECKVYSDLDLYLGVQAINWTEMFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRDAPDIRVNVWLHKETASIALDLSGDGLH
                     LRGYRDRAGIAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMLATDRAPGL
                     HRGRWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIQRARTNAR
                     LAGIGELITFEVKDVAQLTNPLPKEPYGTVLSNPPYGERLDSEPALIALHSLLGRIMK
                     NQFGGWNLSLFSASPDLLSCLQLRADKQYKAKNGPLDCVQKNYHVAESTPDSKPAMVA
                     EDYANRLRKNLKKFEKWARQEGIECYRLYDADLPEYNVAVDRYADWVVVQEYAPPKTI
                     DAHKARQRLFDIIAATISVLGIAPNKLVLKTRERQKGKNQYQKLGEKGEFLEVTEYNA
                     HLWVNLTDYLDTGLFLDHRIARRMLGQMSKGKDFLNLFSYTGSATVHAGLGGARSTTT
                     VDMSRTYLEWAERNLRLNGLTGRAHRLIQADCLAWLREANEQFDLIFIDPPTFSNSKR
                     MEDAFDVQRDHLALMKDLKRLLRAGGTIMFSNNKRGFRMDLDGLAKLGLKAQEITQKT
                     LSQDFARNRQIHNCWLITAA"
     gene            2270. .4177
     CDS             2270. .4177
                     /gene="ECs1033"
                     /note="similar to UUP_ECOLI gi|1787182 percent identity 99
                     in 635 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative ATP-binding component of a transport
                     system"
                     /protein_id="BAB34456.1"
                     /db_xref="GI:13360493"
                     /translation="MSLISMHGAWLSFSDAPLLDNAELHIEDNERVCLVGRNGAGKST
                     LMKILNREQGLDDGRIIYEQDLIVARLQQDPPRNVEGSVYDFVAEGIEEQAEYLKRYH
                     DISRLVMNDPSEKNLNELAKVQEQLDHHNLWQLENRINEVLAQLGLDPNVALSSLSGG
                     WLRKAALGRALVSNPRVLLLDEPTNHLDIETIDWLEGFLKTFNGTIIFISHDRSFIRN
                     MATRIVDLDRGKLVTYPGNYDQYLLEKEEALRVEELQNAEFDRKLAQEEVWIRQGIKA
                     RRTRNEGRVRALKAMRRERGERREVMGTAKMQVEEASRSGKIVFEMEDVCYQVDGKQL
                     VKDFSAQVLRGDKIALIGPNGCGKTTLLKLMLGQLQADSGRIHVGTKLEVAYFDQHRA
                     ELDPDKTVMDNLAEGKQEVMVNGKPRHVLGYLQDFLFHPKRAMTPVRALSGGERNRLL
                     LARLFLKPSNLLILDEPTNDLDVETLELLEELIDSYQGTVLLVSHDRQFVDNTVTECW
                     IFEGGGKIGRYVGGYHDARGQQEQYVALKQPAVKKNEEPAAPKAETVKRSSSKLSYKL
                     QRELEQLPQLLEDLEAKLEALQTQVADASFFSQPHEQTQKVLADMAAAEQELEQAFER
                     WEYLEALKNGG"
     gene            4307. .5560
     CDS             4307. .5560
                     /gene="ECs1034"
                     /note="similar to PQIA_ECOLI gi|1787183 percent identity
                     99 in 417 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="paraquat-inducible protein A"
                     /protein_id="BAB34457.1"
                     /db_xref="GI:13360494"
                     /translation="MCEHHHAAKHILCSQCDMLVALPRLEHGQKAACPRCGTTLTVAW
                     DAPRQRPTAYALAALFMLLLSNLFPFVNMNVAGVTSEITLLEIPGVLFSEDYASLGTF
                     FLLFVQLVPAFCLITILLLVNRAELPVRLKEQLARVLFQLKTWGMAEIFLAGVLVSFV
                     KLMAYGSIGVGSSFLPWCLFCVLQLRAFQCVDRRWLWDDIAPMPELRQPLKPGVTGIR
                     QGLRSCSCCTAILPADEPVCPRCGTKGYVRRRNSLQWTLALLVTSIMLYLPANILPIM
                     VTDLLGSKMPSTILAGVILLWSEGSYPVAAVIFLASIMVPTLKMIAIAWLCWDAKGHG
                     KRDSERMHLIYEVVEFVGRWSMIDVFVIAVLSALVRMGGLMSIYPAMGALMFALVVIM
                     TMFSAMTFDPRLSWDRQPESEHEES"
     gene            5565. .7205
     CDS             5565. .7205
                     /gene="ECs1035"
                     /note="similar to PQIB_ECOLI gi|1787184 percent identity
                     99 in 546 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="paraquat-inducible protein B"
                     /protein_id="BAB34458.1"
                     /db_xref="GI:13360495"
                     /translation="MESNNGEAKIQKVKNWSPVWIFPIVTALIGAWVLFYHYSHQGPE
                     VTLITANAEGIEGGKTTIKSRSVDVGVVESATLADDLTHVEIKARLNSGMEKLLHKDT
                     VFWVVKPQIGREGISGLGTLLSGVYIELQPGAKGSKMDKYDLLDSPPLAPPDAKGIRV
                     VLDSKKAGQLSPGDPVLFRGYRVGSVETSTFDTQKRNISYQLFINAPYDRLVTSNVRF
                     WKDSGIAVDLTSAGMRVEMGSLTTLLSGGVSFDVPEGLDLGQPVAPKTAFVLYDDQKS
                     IQDSLYTDHIDYLMFFKDSVRGLQPGAPVEFRGIRLGTVSKVPFFAPNMRQTFNDDYR
                     IPVLIRIEPERLKMQLGENADVVEHLGELLKRGLRGSLKTGNLVTGALYVDLDFYPNT
                     PAITGIREFNGYQIIPTVSGGLAQIQQRLMEALDKINKLPLNPMIEQATSTLSESQRT
                     MKNLQTTLDSMNKILASQSMQQLPTDMQSTLRELNRSMQGFQPGSAAYNKMVADMQRL
```

```
                DQVLRELQPVLKTLNEKSNALVFEAKDKKDPEPKRAKQ"
     gene       7217. .7765
                /gene="ECs1036"
     CDS        7217. .7765
                /gene="ECs1036"
                /note="similar to YMBA_ECOLI gi|1787185 percent identity
                98 in 182 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB34459.1"
                /db_xref="GI:13360496"
                /translation="MTIAALWLAGCSSGEINKNYYQLPVVQSGTQSTASQGNRLLWVE
                QVAVPDYLAGNGVVYQTSDVKYVIANNNLWASPLDQQLRNTLVANLSTQLPGWVVASQ
                PLGSAQDTLNVTVTEFNGRYDGKVIVSGEWLLNHQGQLIKRPFRLEGVQTQDGYDEMV
                KVLAGVWSQEAASIAQEIKRLP"
     gene       8021. .8188
                /gene="ECs1037"
     CDS        8021. .8188
                /gene="ECs1037"
                /note="identical to RMF_ECOLI gi|1787186 (Conserved in
                E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="ribosome modulation factor"
                /protein_id="BAB34460.1"
                /db_xref="GI:13360497"
                /translation="MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGW
                REAMADRVVMA"
     gene       complement(8258. .8776)
                /gene="ECs1038"
     CDS        complement(8258. .8776)
                /gene="ECs1038"
                /note="similar to FABA_ECOLI gi|1787187 percent identity
                100 in 172 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11

  Query Match          86.6%;  Score 395;  DB 1;  Length 327773;
  Best Local Similarity  92.8%;  Pred. No. 1.3e-103;
  Matches  426;  Conservative    0;  Mismatches   30;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     326421 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 326480

Qy         61 GGTGTTGTTCCTCAGTACGGCGGC---GGCGGTAACCACGGTGGTGGCGGTAATAATAGC 117
              |||||||||||||||||||||||   |||||||||||||||||||||||||||||| |||
Db     326481 GGTGTTGTTCCTCAGTACGGCGGCGGTGGCGGTAACCACGGTGGTGGCGGTAATAACAGC 326540

Qy        118 GGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTG 177
              ||||| ||||| |||||||| ||||| |||||||||| ||||||||||||||||||||
Db     326541 GGCCCGAATTCAGAGCTGAATATTTATCAGTACGGTGGTGGTAACTCTGCACTTGCTCTG 326600

Qy        178 CAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCA 237
              ||| ||||||| |||||||||| | |||||||||||||||| || ||||| ||||||
Db     326601 CAAGCTGATGCTCGTAACTCTGATCTTACTATTACCCAGCATGGTGGTGGTAACGGTGCA 326660

Qy        238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAAC 297
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     326661 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTTGGTAAC 326720

Qy        298 AGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGT 357
              ||||| |||||||||||||||||| ||| | ||| | ||||| |||||||| ||||||
Db     326721 AGCGCCACTCTTGATCAGTGGAACGGTAAAGACTCTCATATGACAGTTAAACAATTCGGT 326780

Qy        358 GGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAG 417
              || |||||||||| || |||||||||||||||||||| ||| |||||||| ||||||
Db     326781 GGCGGCAACGGTGCAGCGGTTGACCAGACTGCATCTAATTCCACCGTCAACGTAACTCAG 326840

Qy        418 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||
Db     326841 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 326879
```

```
RESULT 9
AE016759
LOCUS       AE016759               306358 bp    DNA     linear   BCT 24-DEC-2002
DEFINITION  Escherichia coli CFT073 section 5 of 18 of the complete genome.
ACCESSION   AE016759 AE014075
VERSION     AE016759.1  GI:26107527
KEYWORDS    .
SOURCE      Escherichia coli CFT073
  ORGANISM  Escherichia coli CFT073
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Extensive Mosiac Structure Revealed by the Complete Genome Sequence
            of Uropathogenic Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
   PUBMED   12471157
REFERENCE   2  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .306358
                     /organism="Escherichia coli CFT073"
                     /mol_type="genomic DNA"
                     /strain="CFT073"
                     /db_xref="taxon:199310"
     gene            complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
     CDS             complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
                     /note="Residues 27 to 1231 of 1245 are 80.16 pct identical
                     to residues  1 to 1209 of 1218 from GenPept.129 :
                     >emb|CAD05883.1| (AL627276) putative ABC transporter
                     protein [Salmonella enterica subsp. enterica serovar
                     Typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP binding cassette (ABC) transporter homolog"
                     /protein_id="AAN79727.1"
                     /db_xref="GI:26107528"
                     /translation="MRDRSVLAFLRNDPASGSYFTDGGGLMPANHTPTPAQSWIVRLA
                     RVCWERKKLSVIVVVASVSTILLAALTPLLTRQAVNDALAGNPARLPWLACGLLLIAF
                     FDFIGNYVRRGYAGMLSLWVQHTLRGRVFDSIQKLDGAGQDALRTGQVISRTNSDLQQ
                     VHTLLQMCPVPLAVFTYYIAGIAVMLWMSPAMTLIVVCVLVCLAITALRARRRVFAQT
                     GMASDQLANLTEHIREVLAQISVVKSCVAEMRETHWLDRQSRQIVRVRIGAVISQAMP
                     GATMLALPVLGQIVLLCYGGWSVMHGRIDLGTFVAFASFLAMLTGPTRVLASFLVIAQ
                     RTQASVERVFALIDTRSQMEDGTESINSQVVGLELENMSFDYHHGDRHILSDISFSLR
                     AGETVAVVGASGSGKSTLLMLLARFYDPCSGKIWLNTSEGRQNLRDIRLEALRRRVGI
                     VFEDAFLFAGTVAENIAYGHPQATADDIRRAAAAAGASDFINALPKGFDSLLTERGTN
                     LSGGQRQRIALARALITAPDVLILDDTTSAVDAVTEAEINTALGRYADEGHMLLVIAR
                     RRSTLQLASRVVVLDKGRMVDTGTPAELEARCPAFRALMTGDSDFLATSHNSHNELWP
                     AEPATQDDVTDTGDKGFVARMTRVPENAVQQALAGKGRKVTSLLKPVAWMFVIAALLI
                     ALDSAAGVGVLILLQHGIDSGVAAGDMSTIGLCALLALCLVIVGWCSYSLQTVFAARA
                     ABSVQHSVRLRSFGHMLRLGLPWHEKHADSRLTRMTVDVDSLARFLQNGLAGAATSLV
                     TMFAIAATMFWLDPILALTALSAVPVAALATMIYRRLSTPAYAQARLEIGKVNSTLQE
                     KVSGLRVVQSHGQQELEGARLRALSERFRATRVRAQKYLAVYFPFLTFCTEASYAAVL
                     LVGASQVAAGEMTAGVLAAFFLLLGQFYGPVQQLSGIVDAWQQATASGKHIDELLATE
                     GTENLGSSSVLPVTGALHLDEVTFSYPDSHEPALNKLTLTIPEGMVVAVVGRSGAGKS
```

```
                TLIKLIAGLYFPTHGNIRIGVQMLDDASLTEYRRQIGLVDQDVALFSSDIAENIRYSR
                PSATNEDVEIASQRAGLYEMVCNLPQGFRTPVNNGGADLSAGQRQLIALARAQLANAH
                ILLLDEATSCLDRTSEERLMSSLTDVVHAGKHSALIVAHRLTTAQRCDLIAVIDKGLL
                AEYGTHEQLLSAGGLYTRLWHDSVSSTALHRQHNMKEETPG"
gene            complement(3882. .5045)
                /gene="iroB"
                /locus_tag="c1254"
CDS             complement(3882. .5045)
                /gene="iroB"
                /locus_tag="c1254"
                /note="Escherichia coli O157:H7 ortholog: z1190"
                /codon_start=1
                /transl_table=11
                /product="Putative glucosyltransferase"
                /protein_id="AAN79728.1"
                /db_xref="GI:26107529"
                /translation="MRRLPDLDRQAERDFLMRILFVGPPLYGLLYPVLSLAQAFRVNG
                HEVLIASGGQFAQKAAEAGLVVFDAAPGLDSEAGYRHHEAQRKKSNIGTQMGNFSFFS
                EEMADHLVEFAGHWRPDLIIYPPLGVIGPLIAAKYDIPVVMQTVGFGHTPWHIRGVTR
                SLTDAYRRHNVGATPRDMAWIDVTPPSMSILENDGEPIIPMQYVPYNGGAVWEPWWER
                RPDRKRLLVSLGTVKPMVDGLDLIAWVMDSASEVDAEIILHISANARSDLRSLPSNVR
                LVDWIPMGVFLNGADGFIHHGGAGNTLTALHAGIPQIVFGQGADRPVNARVVAERGCG
                IIPGDVGLSSNMINAFLNNRSLRKASEEVAAEMAAQPCPGEVAKSLITMVQKG"
gene            5535. .5834
                /locus_tag="c1255"
CDS             5535. .5834
                /locus_tag="c1255"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79729.1"
                /db_xref="GI:26107530"
                /translation="MKPTLLQTVENICLQILFVLFLFTGTLIDVEALVSCSSVAKNVQ
                RTPASSVSSSVYSASVSRTAADIMLLSNNPLFAPGQVSFFGCSLAASGIHVTILE"
repeat_region   5867. .6577
                /insertion_seq="IS1541A-like"
gene            5961. .6470
                /locus_tag="c1256"
CDS             5961. .6470
                /locus_tag="c1256"
                /note="Residues 18 to 169 of 169 are 94.07 pct identical
                to residues  1 to 152 of 152 from SwissProt.40 :
                >sp|Q57334|T200_SALTY Transposase for insertion sequence
                element IS200"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79730.1"
                /db_xref="GI:26107531"
                /translation="MRSGNCKCQTRNQKGVPMGNEKSLAHTRWNCKYHIVFAPKYRRQ
                VFYREKRRAIGCILRKLCEWKSVRILEAECCADHIHMLVEIPPKMSVSGFMGYLKGKS
                SLMPYEQFGDLKFKYRNREFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPY
                PGSPFTGRK"
repeat_region   complement(6578. .6940)
                /note="incomplete"
                /insertion_seq="IS4"
gene            complement(6598. .6960)
                /locus_tag="c1257"
CDS             complement(6598. .6960)
                /locus_tag="c1257"
                /function="IS, phage, Tn"
                /note="Residues 8 to 120 of 120 are 98.23 pct identical to
                residues 316 to 428 of 442 from MG1655 : b4278"
                /codon_start=1
                /transl_table=11
                /product="Putative conserved protein"
                /protein_id="AAN79731.1"
                /db_xref="GI:26107532"
                /translation="MPRQPDQGYREIKQTMQLSRLTLRSKKPELVEQELWGVLLAYNL
                VRYQMIKMAEHLKGYWPNQLSFSESCGMVMRMLMTLQGASPGRIPELMRDLASMGQLV
                KLPTRRETAFPRVVKERP"
gene            7651. .8013
                /locus_tag="c1258"
CDS             7651. .8013
                /locus_tag="c1258"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79732.1"
                /db_xref="GI:26107533"
                /translation="MLGACFYSWGTSFYYGFPWWGAELVLMMWPEACFMLLQSSLYSL
                LDGEQVLFSFWELNKKQYEGFEFYQIFSGNIVASCSSCSGVSSNSSAHYARCTDFMRH
                YSLNNYAACQLGRQACLS"
gene            7835. .8269
                /locus_tag="c1259"
CDS             7835. .8269
                /locus_tag="c1259"
                /note="Residues 1 to 139 of 144 are 50.71 pct identical to
                residues  104 to 242 of 258 from GenPept.129 :
                >gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
                [Shigella flexneri]"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79733.1"
                /db_xref="GI:26107534"
                /translation="MKDLSFIRFFLAILLLLVPPALEFPVIHRHITPGALTLCVITAL
                IITLLASLGGRLVSVKCISEISFIRRHRIECMMAGFMIYFWTFSLIAGWYKPQFKKEY
                QTLHYENARYYVLARYDGRLVLSQSYRSGSRKFVIINGGHRL"
gene            8477. .8553
                /locus_tag="c5534"
tRNA            8477. .8553
                /locus_tag="c5534"
                /product="tRNA-Arg"
                /function="RNA; tRNA"
gene            complement(8832. .9290)
                /locus_tag="c1260"
CDS             complement(8832. .9290)
                /locus_tag="c1260"
                /function="putative enzyme; Transposases Insertion
                Sequence Associated"
                /note="Escherichia coli O157:H7 ortholog: z4324"
                /codon_start=1
                /transl_table=11
                /product="Putative Transposase"
                /protein_id="AAN79734.1"
                /db_xref="GI:26107535"
                /translation="MASAADARRIVSHYERRWLIEEYHKAWKSGGTCVESLRMQTRDN
                LERMVVIKAFIAVRMLGLRQEGISEETQNDSCKKILTPTEWKLLWVKLEGKQLPSQTP
                TLKWACLKLGRWHDSKRTGRPGWVVMWDGWFRLQDMVEGYPVMKSLDQEI"
gene            complement(9302. .9667)
                /locus_tag="c1261"
CDS             complement(9302. .9667)
                /locus_tag="c1261"
                /note="Residues 1 to 35 of 121 are 97.14 pct identical to
                residues  59 to 93 of 316 from GenPept.129 :
                >gb|AAK18437.1|AF348706_126 (AF348706) putative
                transposase [Shigella flexneri]"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /protein_id="AAN79735.1"
                /db_xref="GI:26107536"
                /translation="MAERLGEIQKRVITVCDREADIWHYLHYKVSHGQRFVVRTAQNS
```

```
 Query Match             83.8%;  Score 382.2;  DB 1;  Length 306358;
 Best Local Similarity   91.1%;  Pred. No. 7.1e-100;
 Matches  418;  Conservative    0;  Mismatches   38;  Indels    3;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db      45269 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 45328

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGT---AACCACGGTGGTGGCGGTAATAATAGC 117
              |||||||||||||||||||||||||||||   |||||||||||||||||||||| |||
Db      45329 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTGGCAACCACGGTGGTGGCGGTAATAACAGC 45388
```

```
Qy        118 GGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTG 177
              || || ||||| |||||||| ||||||||||||||||||||||||||||| ||||||| |
Db      45389 GGTCCGAATTCAGAGCTGAATATTTACCAGTACGGTGGCGGTAACTCTGCTCTTGCTCAG 45448

Qy        178 CAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCA 237
              ||| |||| |||||||||||||  |||| |||||||||| |||||||||||||| |||
Db      45449 CAAGCTGACGCCCGTAACTCTGATCTGACCATTACCCAGCACGGCGGCGGTAATGGCGCA 45508

Qy        238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAAC 297
              ||||| || || || || ||||||||||||||||||||||| || ||||| ||||| |||
Db      45509 GATGTGGGCCAAGGTTCTGATGACAGCTCAATCGATCTGACTCAGCGTGGTTTCGGCAAC 45568

Qy        298 AGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGT 357
              |||||||||||||||||||||| || ||| |||||   ||||| ||||||||||||||||
Db      45569 AGCGCTACTCTTGATCAGTGGAATGGTAAAGATTCTACTATGACTGTTAAACAGTTCGGT 45628

Qy        358 GGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAG 417
              || || |||||||||| |||||||||||||||||||||||  ||| ||||| ||||||
Db      45629 GGCGGTAACGGTGCTGCTGTTGACCAGACTGCATCTAACTCCAGCGTTAACGTCACTCAG 45688

Qy        418 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||
Db      45689 GTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 45727
```

```
RESULT 10
AE015131
LOCUS       AE015131               10370 bp    DNA     linear   BCT 18-OCT-2002
DEFINITION  Shigella flexneri 2a str. 301 section 94 of 412 of the complete
            genome.
ACCESSION   AE015131 AE005674
VERSION     AE015131.1  GI:24051313
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 301
  ORGANISM  Shigella flexneri 2a str. 301
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 10370)
  AUTHORS   Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
            Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
            Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
            Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
            Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
  TITLE     Genome sequence of Shigella flexneri 2a: insights into
            pathogenicity through comparison with genomes of Escherichia coli
            K12 and O157
  JOURNAL   Nucleic Acids Res. 30 (20), 4432-4441 (2002)
   PUBMED   12384590
REFERENCE   2  (bases 1 to 10370)
  AUTHORS   Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B.,
            Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y.,
            Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y.,
            Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
            of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.
            China
FEATURES             Location/Qualifiers
     source          1. .10370
                     /organism="Shigella flexneri 2a str. 301"
                     /mol_type="genomic DNA"
                     /strain="301"
                     /serotype="2a"
                     /db_xref="taxon:198214"
     tRNA            complement(127. .214)
                     /product="tRNA-Ser"
                     /note="anticodon: GGA"
     gene            410. .1387
                     /gene="ycdW"
                     /locus_tag="SF1027"
                     /pseudo
     CDS             410. .1387
                     /gene="ycdW"
                     /locus_tag="SF1027"
                     /note="similar to Escherichia coli K12 putative
                     dehydrogenase gi: 1787270 (326 aa). BLAST with identity of
                     99% in 326 aa. This CDS contains an in-frame stop codon.
                     The sequence has been checked and is believed to be
                     correct."
                     /pseudo
                     /codon_start=1
                     /transl_table=11
     gene            1442. .2179
                     /gene="ycdX"
                     /locus_tag="SF1028"
     CDS             1442. .2179
                     /gene="ycdX"
                     /locus_tag="SF1028"
                     /note="Residues 1 to 245 of 245 are 99 pct identical to
                     residues 1 to 245 of a 245 aa protein from Escherichia
                     coli O157:H7 EDL933 ref: NP_287168.1"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, conserved hypothetical protein"
                     /protein_id="AAN42652.1"
                     /db_xref="GI:24051314"
                     /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                     EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                     PVFAPHDKATNTQAMISTIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                     SFLHSRKGSEDNCRAVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPLER
                     ILNVSPRRLLNFLESRGMAPIAEFADL"
     gene            2203. .2757
                     /gene="ycdY"
                     /locus_tag="SF1029"
     CDS             2203. .2757
                     /gene="ycdY"
                     /locus_tag="SF1029"
                     /note="Residues 1 to 184 of 184 are 99 pct identical to
                     residues 1 to 184 of a 184 aa protein from Escherichia
                     coli O157:H7 EDL933 ref: NP_287169.1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative oxidoreductase component"
                     /protein_id="AAN42653.1"
                     /db_xref="GI:24051315"
                     /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                     DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEDATEAEVRAFLSERGMLL
                     ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                     RTMAPLTRDAISAMWDELEEDSEE"
     gene            2811. .3350
                     /gene="ycdZ"
                     /locus_tag="SF1030"
     CDS             2811. .3350
                     /gene="ycdZ"
                     /locus_tag="SF1030"
                     /note="Residues 1 to 179 of 179 are 98 pct identical to
                     residues 1 to 179 of a 179 aa protein from Escherichia
                     coli K12 ref: NP_415554:1"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, conserved hypothetical protein"
                     /protein_id="AAN42654.1"
                     /db_xref="GI:24051316"
                     /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                     SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                     IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLVFGYAMKNSGLW
                     LAARSAKTAHREQEIKNKA"
     gene            complement(3414. .4189)
                     /gene="csgG"
                     /locus_tag="SF1031"
                     /pseudo
     CDS             complement(3414. .4189)
                     /gene="csgG"
                     /locus_tag="SF1031"
                     /note="similar to Escherichia coli K12 curli production
```

assembly-transport component, 2nd curli operon gi: 1787274 (278 aa). BLAST with identity of 97% in 257 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11

gene complement(4454. .5662)

CDS complement(4454. .5662)
/locus_tag="SF1032"
/note="Residues 1 to 402 of 402 are 100 pct identical to residues 1 to 402 of a 402 aa protein from Plasmid R100 ref: NP_052925.1"
/codon_start=1
/transl_table=11
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translation="MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTLTLTEL GRNLPTKARTKHNIKRIDRLLGNRHLHKERLAVYRWHASFICSGNTMPIVLVDWSDIR EQKRLMVLRASVALHGRSVTLYEKAFPLSEQCSKKAHDQFLADLASILPSNTTPLIVS DAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAENWKPISNLHDMSSSHSKTLGYKR LTKSNPISCQILLYKSRSKGRKNQRSTRTHCHHPSPKIYSASAKEPWILATNLPVEIR TPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWL AGVHAQKQGWDKHFQANTVRNRNVLSTVRLGMEVLRHSGYTITREDSLVAATLLTQNL FTHGYVLGKL"

gene complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"

CDS complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"
/note="Residues 1 to 129 of 129 are 98 pct identical to residues 1 to 129 of a 129 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287173.1"
/codon_start=1
/transl_table=11
/product="curli production assembly-transport component, 2nd curli operon"
/protein_id="AAN42656.1"
/db_xref="GI:24051318"
/translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDSEKTVVFAL IQTEEALNRRQINQALLSTDDLAHDEF"

gene complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"

CDS complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"
/note="Residues 11 to 226 of 226 are 100 pct identical to residues 1 to 216 of a 216 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=11
/product="putative 2-component transcriptional regulator for 2nd curli operon"
/protein_id="AAN42657.1"
/db_xref="GI:24051319"
/translation="MRSIKKWGFIMFNEVHSIHGHTLLLITKPSLQATALLQHLKQSL AITGKLHNIQRSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPED YPYRDIENWPHINGVFYAMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNS TESALLTHREKEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVS WANDNLRR"

gene 7766. .8248
/gene="csgB"

CDS 7766. .8248
/gene="csgB"
/locus_tag="SF1035"
/note="Residues 10 to 160 of 160 are 99 pct identical to residues 1 to 151 of a 151 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287175.1"
/codon_start=1
/transl_table=11
/product="minor curlin subunit precursor, similar ro CsgA"
/protein_id="AAN42658.1"
/db_xref="GI:24051320"
/translation="MYDQVQGDNMKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAV NELSKSSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYI DQAGSANDASISQGAYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR

Query Match 74.7%; Score 340.8; DB 1; Length 10370;
Best Local Similarity 98.0%; Pred. No. 5.2e-88;
Matches 345; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy      105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
            ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     9538 CAGTAATAATAGCGGCCCAAATTCAGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 9597

Qy      165 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 224
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9598 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 9657

Qy      225 CGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 284
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9658 TGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 9717

Qy      285 TGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGT 344
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||| ||
Db     9718 TGGCTTCGGTAACAGCGCCACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACTGT 9777

Qy      345 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGT 404
            |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     9778 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCGAACTCCTCCGT 9837

Qy      405 CAACGTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
            |||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     9838 CAACGTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTATTAA 9889
```

RESULT 11
AE016981
LOCUS AE016981 292504 bp DNA linear BCT 22-APR-2003
DEFINITION Shigella flexneri 2a str. 2457T section 4 of 16 of the complete genome.
ACCESSION AE016981 AE014073
VERSION AE016981.1 GI:30040616
KEYWORDS .
SOURCE Shigella flexneri 2a str. 2457T
ORGANISM Shigella flexneri 2a str. 2457T
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 292504)
AUTHORS Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W., Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A., Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S., Schwartz,D.C. and Blattner,F.R.
TITLE Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T
JOURNAL Infect. Immun. 71 (5), 2775-2786 (2003)
PUBMED 12704152
REFERENCE 2 (bases 1 to 292504)
AUTHORS Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W., Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A., Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S., Schwartz,D.C. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES Location/Qualifiers
source 1. .292504
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"

```
                /serotype="2a"
                /db_xref="taxon:198215"
     gene       complement(223. .441)
                /gene="infA"
                /locus_tag="S0884"
     CDS        complement(223. .441)
                /gene="infA"
                /locus_tag="S0884"
                /function="factor; Proteins - translation and
                modification"
                /note="residues 1 to 72 of 72 are 100.00 pct identical to
                residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
                /codon_start=1
                /transl_table=11
                /product="protein chain initiation factor IF-1"
                /protein_id="AAP16348.1"
                /db_xref="GI:30040617"
                /translation="MAKEDNIEMQGTVLETLPNTMFRVELENGHVVTAHISGKMRKNY
                IRILTGDKVTVELTPYDLSKGRIVFRSR"
     gene       complement(726. .1430)
                /gene="aat"
                /locus_tag="S0885"
     CDS        complement(726. .1430)
                /gene="aat"
                /locus_tag="S0885"
                /function="enzyme; Aminoacyl tRNA synthetases, tRNA
                modification"
                /note="residues 1 to 234 of 234 are 100.00 pct identical
                to residues 1 to 234 of 234 from Escherichia coli K-12 :
                B0885"
                /codon_start=1
                /transl_table=11
                /product="leucyl, phenylalanyl-tRNA-protein transferase"
                /protein_id="AAP16349.1"
                /db_xref="GI:30040618"
                /translation="MRLVQLSRHSIAFPSPEGALREPNGLLALGGDLSPARLLMAYQR
                GIFPWFSPGDPILWWSPDPRAVLWPESLHISRSMKRFHKRSPYRVTMNYAFGQVIEGC
                ASDREEGTWITRGVVEAYHRLHELGHAHSIEVWREDELVGGMYGVAQGTLFCGESMFS
                RMENASKTALLVFCEEFIGHGGKLIDCQVLNDHTASLGACEIPRRDYLNYLNQMRLGR
                LPNNFWVPRCLFSPQE"
     gene       complement(1472. .3193)
                /gene="cydC"
                /locus_tag="S0886"
     CDS        complement(1472. .3193)
                /gene="cydC"
                /locus_tag="S0886"
                /function="transport; Transport of small molecules: Other"
                /note="residues 1 to 573 of 573 are 97.03 pct identical to
                residues 1 to 573 of 573 from Escherichia coli K-12 :
                B0886"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding component of cytochrome-related
                transport"
                /protein_id="AAP16350.1"
                /db_xref="GI:30040619"
                /translation="MRALLPYLALYKRHKWMLSLGIVLAIVTLLASIGLLTLSGWFLS
                ASAVAGVAGLYSFNYMLPAAGVRGAAITRTAGRYFERLVSHDATFRVLQHLRIYTFSK
                LLPLSPAGLARYRQGELLNRVVADVDTLDHLYLRVISPLVGAFVVIMVVTIGLSFLDF
                TLAFTLGSVMLLTLFLMPPLFYRAGKSTGQNLTHLRGQYRQQLTAWLQGQAELTIFGA
                SDRYRTQLENTEIQWLEAQRRQSELTALSQAIMLLIGALAVILMLWMASGGVGGNAQP
                GALIALFVFCALAAFEALAPVTGAFQHLGQVIASAVRISDLTDQKPEVTFPDTQTRVA
                DRVSLTLRDVQFTYPEQSQQALKGISLQVNAGEHIAILGRTGCGKSTLLQLLTRAWDP
                QQGEILLNDSPIASLNEAALRQTISVVPQRVHLFSATLRDNLLLASPGSSDEALSEIL
                RRVGLEKLLEDAGLNSWLGEGGRQLSGGELRRLAIARTLLHDAPLVLLDEPTEGLDAT
                TESQILELLAKMMREKTVLMVTHRLRGLSRFQQIIVMDNGQIIEQGTHAELLARQGRY
                YQFKQGL"
     gene       complement(3194. .4960)
                /gene="cydD"
                /locus_tag="S0887"
     CDS        complement(3194. .4960)
                /gene="cydD"
                /locus_tag="S0887"
                /function="transport; Not classified"
                /note="residues 1 to 588 of 588 are 92.85 pct identical to
                residues 1 to 588 of 588 from Escherichia coli K-12 :
                B0887"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding component of cytochrome-related
                transport"
                /protein_id="AAP16351.1"
                /db_xref="GI:30040620"
                /translation="MNKSRQKELTRWLKQQSVISQRWLNISRLLGFVSGILIIAQAWF
                MARILQHMIMENIPREALLLPFTLLVLTFVLRAWVVWLRERVGYHAGQHIRFAIRRQV
                LDRLQQAGPAWIQGKPAGSWATLVLEQIDDMHDYYARYLPQMALAVSVPLLIVVAIFP
                SNWAAALILLGTAPLIPLFMALVGMGAADANRRNFLALARLSGHFLDRLRGMETLRIF
                GRGEAEIESIRSASEDFRQRTMEVLRLAFLSSGILEFFTSLSIALVAVYFGFSYLGEL
                DFGHYDTGVTLAAGFLALILAPEFFQPLRDLGTFYHAKAQAVGAADSLKTFMETPLAH
                PQRGEAELALTDPLTIEAEDLFITSPEGKTLAGPLNFTLPAGQRAVLVGRSGSGKSSL
                LNALSGFLSYQGSLRINGIELRDLSPESWRKHLSWVGQNPQLPAAKLRDNVLLARPDA
                SEQELQAALDNAWVSEFLPLLPQGVDTPVGDQAARLSVGQAQRVAVARALLNPCSLLL
                LDEPAASLDAHSEQRVMEALNAASLRQTTLMVTHQLEDLADWDVIWVMQDGRIIEQGR
                YAELSVAGGPFATLLAHRQEEI"
     gene       complement(5083. .6048)
                /gene="trxB"
                /locus_tag="S0888"
     CDS        complement(5083. .6048)
                /gene="trxB"
                /locus_tag="S0888"
                /function="enzyme; 2-Deoxyribonucleotide metabolism"
                /note="residues 1 to 321 of 321 are 95.95 pct identical to
                residues 1 to 321 of 321 from Escherichia coli K-12 :
                B0888"
                /codon_start=1
                /transl_table=11
                /product="thioredoxin reductase"
                /protein_id="AAP16352.1"
                /db_xref="GI:30040621"
                /translation="MGTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQL
                TTTTEVENWPGDPNDLTGPLLMARMHEHATKFETEIIFDHINKVDLQNRPFRLNGDNG
                EYTCDALIIATGASARYLGLPSEEAFKGRGVSACATCDGFFYRNQKVAVIGGGNTAVE
                EALYLSNIASEVHLIHRRDGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGV
                TGVRLRDTQNSDNIESLDVAGLFVAIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQ
                TSIPGVFAAGDVMDHIYRQAITSAGTGCMAALDAERYLDGLADAK"
     gene       6592. .7086
                /gene="lrp"
                /locus_tag="S0889"
     CDS        6592. .7086
                /gene="lrp"
                /locus_tag="S0889"
                /function="regulator; Central intermediary metabolism:
                Pool, multipurpose conversions"
                /note="residues 1 to 164 of 164 are 100.00 pct identical
                to residues 1 to 164 of 164 from Escherichia coli K-12 :
                B0889"
                /codon_start=1
                /transl_table=11
                /product="regulator for leucine (or lrp) regulon and
                high-affinity branched-chain amino acid transport system"
                /protein_id="AAP16353.1"
                /db_xref="GI:30040622"
                /translation="MVDSKKRPGKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTP
                CLERVRRLERQGFIQGYTALLNPHYLDASLLVFVEITLNRGAPDVFEQFNTAVQKLEE
                IQECHLVSGDFDYLLKTRVPDMSAYRKLLGETLLRLPGVNDTRTYVVMEEVKQSNRLV
                IKTR"
     gene       7221. .11249
                /gene="ftsK"
                /locus_tag="S0890"
     CDS        7221. .11249
                /gene="ftsK"
                /locus_tag="S0890"
                /function="phenotype; Cell division"
                /note="residues 1 to 737 of 1342 are 68.24 pct identical
                to residues 1 to 737 of 1329 from Escherichia coli K-12 :
                B0890"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein"
                     /protein_id="AAP16354.1"
                     /db_xref="GI:30040623"
                     /translation="MSQEYTEDKEVTLTKLSSGRRLLEALLILIVLFAVWLMAALLSF
                     NPSDPSWSQTAWHEPIHNLGGMPGAWLADTLFFIFGVMAYTIPVIIVGGCWFAWRHQS
                     SDEYIDYFAVSLRIIGVLALILTSCGLAAINADDIWYFASGGVIGSLLSTTLQPLLHS
                     SGGTIALLCVWAAGLTLFTGWSWVTIAEKLGGWILNILTFASNRTRRDDTWVDEDEYE
                     DDEEYEDENHGKQHESRRARILRGALARRKRLAEKFINPMGRQTDAALFSGKRMDDDE
                     EITYTARGVAADPDDVLFSGNRATQPEYDEYDPLLNSAPITEPVAVAAAATTATQSWA
                     APVEPVTQTPPVASVDVPPSQPTVAWQPVPGPQTGEPVIAPAPEGYPQQSQYAQPAVQ
                     YNEPLQQPVQPQQPYYAPAAEQPAQQPYYAPAPEQPVAGNAWQAEEQQSTFAPLSTYQ
                     TEQTYQQPAAQEPLYQQPQPVEQQPVVEPEPVVEETKPARPPLYYFEEVEEKRARERE
                     QLAAWYQPIPEPVKEPEPIKSSLKAPSVAAVPPVEAAAAVSPLASGVKKATLATGAAA
                     TVAAPVFSLANSGGPRPQVKEGIGPQLPRPKRIRVPTRRELASYGIKLPSQRAAEEKA
                     REAQRNQYDSGDQYNDDEIDAMQQDELARQFAQTQQQRYGEQYQHDVPVNAEDADAAA
                     EAELARQFAQTQQQRYSGEQPAGANPFSLDDFEFSPMKALLDDGPHEPLFTPIVEPVQ
                     QPQQPVAPQQQYQQPQQPVAPQQQYQQPQQQVAPQPQYQQPQQPVAPQQQYQQPQQPV
                     APQPQYQQPQQPVAPQPQYQQPQQPVAPQPQDTLLHPLLMRNGDSRPLHKPTTPLPSL
                     DLLTPPPSEVEPVDTFALEQMARLVEARLADFRIKADVVNYSPGPVITRFELNLAPGV
                     KAARISNLSRDLARSLSTVAVRVVEVIPGKPYVGLELPNKKRQTVYLREVLDNAKFRD
```

Query Match 74.7%; Score 340.8; DB 1; Length 292504;
Best Local Similarity 98.0%; Pred. No. 8.3e-88;
Matches 345; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
              | |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     212560 CAGTAATAATAGCGGCCCAAATTCAGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 212619

Qy        165 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212620 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 212679

Qy        225 CGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212680 TGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 212739

Qy        285 TGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGT 344
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||| ||
Db     212740 TGGCTTCGGTAACAGCGCCACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACTGT 212799

Qy        345 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGT 404
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     212800 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCGAACTCCTCCGT 212859

Qy        405 CAACGTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     212860 CAACGTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTATTAA 212911
```

```
RESULT 12
AF237726/c
LOCUS       AF237726               437 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Shigella flexneri insertion sequence IS600, complete sequence.
ACCESSION   AF237726
VERSION     AF237726.1  GI:7582371
KEYWORDS    .
SOURCE      Shigella flexneri
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Sakellaris,H., Hannink,N.K., Rajakumar,K., Bulach,D., Hunt,M.,
            Sasakawa,C. and Adler,B.
  TITLE     Curli loci of Shigella spp
  JOURNAL   Infect. Immun. 68 (6), 3780-3783 (2000)
  MEDLINE   20278182
   PUBMED   10816548
REFERENCE   2  (bases 1 to 437)
  AUTHORS   Sakellaris,H., Hannink,N.K., Rajakumar,K., Bulach,D., Hunt,M.,
            Sasakawa,C. and Adler,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2000) Microbiology, Monash University, Wellington
            Rd., Clayton, Vic 3800, Australia
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Shigella flexneri"
                     /mol_type="genomic DNA"
                     /strain="SBA1100"
                     /serotype="2a"
                     /db_xref="taxon:623"
     source          320. .392
                     /organism="Shigella flexneri"
                     /mol_type="genomic DNA"
                     /strain="SBA1100"
                     /serotype="2a"
                     /db_xref="taxon:623"
                     /insertion_seq="IS600"
                     /note="similar to IS600 deposited in GenBank Accession
                     Number X05952"
     misc_feature    join(1. .319,393. .437)
                     /note="csg locus; similar to Escherichia coli K12 csg
                     locus deposited in GenBank Accession Number X90754"
ORIGIN
```

Query Match 68.3%; Score 311.4; DB 1; Length 437;
Best Local Similarity 98.1%; Pred. No. 1.2e-79;
Matches 315; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy        105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
              | |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        321 CAGTAATAATAGCGGCCCAAATTCAGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 262

Qy        165 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 TGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGG 202

Qy        225 CGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACG 142

Qy        285 TGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGT 344
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||| ||
Db        141 TGGCTTCGGTAACAGCGCCACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACTGT 82

Qy        345 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGT 404
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         81 TAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCAGACTGCATCGAACTCCTCCGT 22

Qy        405 CAACGTGACTCAGGTTGGCTT 425
              |||||||||||||||||||||
Db         21 CAACGTGACTCAGGTTGGCTT 1
```

```
RESULT 13
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
```

```
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYFSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
     gene            3344. .3376
                     /gene="csgBA"
     -35_signal      3344. .3349
                     /gene="csgBA"
     -10_signal      3371. .3376
                     /gene="csgBA"
     gene            3469. .3924
                     /gene="csgB"
     CDS             3469. .3924
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="CAA05316.1"
                     /db_xref="GI:2739238"
                     /db_xref="GOA:P55226"
                     /db_xref="SWISS-PROT:P55226"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     3469. .3531
                     /gene="csgB"
     mat_peptide     3532. .3921
                     /gene="csgB"
                     /product="CsgB protein"
                     /function="nucleator"
     gene            3966. .4421
                     /gene="csgA"
     CDS             3966. .4421
                     /gene="csgA"
                     /function="fiber subunit"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="CAA05317.1"
                     /db_xref="GI:2739239"
                     /db_xref="GOA:P55225"
                     /db_xref="SWISS-PROT:P55225"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     3966. .4025
                     /gene="csgA"
     mat_peptide     4026. .4418
                     /gene="csgA"
                     /product="CsgA protein"
                     /function="fiber subunit"
     gene            4483. .4809
                     /gene="csgC"
     CDS             4483. .4809
                     /gene="csgC"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="CsgC protein"
                     /protein_id="CAA05318.1"
                     /db_xref="GI:2739240"
                     /db_xref="SWISS-PROT:P55227"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
     sig_peptide     4483. .4533
                     /gene="csgC"
     mat_peptide     4534. .4806
                     /gene="csgC"
                     /product="CsgC"
     gene            4940. .5103
                     /gene="ORF103"
     CDS             4940. .>5103
                     /gene="ORF103"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA05319.1"
                     /db_xref="GI:2739241"
                     /db_xref="SWISS-PROT:O54295"
                     /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                     CDISTHAENI"
ORIGIN

  Query Match           56.8%;  Score 259.2;  DB 1;  Length 5103;
  Best Local Similarity  73.0%;  Pred. No. 2.9e-64;
  Matches  333;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db       3966 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 4025

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db       4026 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 4085

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  ||||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db       4086 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 4145

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              | |||||||||||| |||||   ||| ||||||||||   ||    ||||| || || |||
Db       4146 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 4205

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| |||| | || || ||||| ||   ||| ||| | ||||
Db       4206 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 4265

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || |||||||||  ||||| || || || || ||    || | ||| |||
Db       4266 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 4325

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db       4326 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 4385

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db       4386 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 4421


RESULT 14
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .22411
                     /organism="Salmonella typhimurium LT2"
                     /mol_type="genomic DNA"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            434. .1308
                     /gene="phoH"
                     /note="synonym: STM1126"
     RBS             434. .439
                     /gene="phoH"
                     /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS             454. .1308
                     /gene="phoH"
                     /note="similar to E. coli PhoB-dependent, ATP-binding pho
                     regulon component; may be helicase; induced by P
                     starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                     aa), 92% identity in aa 71 - 354"
                     /codon_start=1
                     /transl_table=11
                     /product="PhoB-dependent ATP-binding pho regulon
                     component"
                     /protein_id="AAL20057.1"
                     /db_xref="GI:16419642"
                     /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                     RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                     GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                     YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
```

```
                LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                QRTLHAYS"
     gene       complement(1414. .2303)
                /gene="STM1127"
     CDS        complement(1414. .2295)
                /gene="STM1127"
                /note="similar to E. coli orf, hypothetical protein
                (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                identity in aa 15 - 264"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /protein_id="AAL20058.1"
                /db_xref="GI:16419643"
                /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                QARESKLRTMNALDMTK"
     RBS        complement(2298. .2303)
                /gene="STM1127"
                /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene       complement(2580. .4088)
                /gene="STM1128"
     CDS        complement(2580. .4076)
                /gene="STM1128"
                /note="similar to E. coli putative cotransporter
                (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                identity in aa 7 - 478"
                /codon_start=1
                /transl_table=11
                /product="putative sodium/glucose cotransporter"
                /protein_id="AAL20059.1"
                /db_xref="GI:16419644"
                /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                DKFFPATQFHWSWTESTVPVLMIGFLFANIQQPTASQDVVQRYIVTDSIEETKKTLLT
                NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS        complement(4083. .4088)
                /gene="STM1128"
                /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene       complement(4413. .5099)
                /gene="STM1129"
     CDS        complement(4413. .5093)
                /gene="STM1129"
                /note="similar to E. coli putative enzyme (AAC76255.1);
                Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                226"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL20060.1"
                /db_xref="GI:16419645"
                /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                WPSHAVKR"
     RBS        complement(5094. .5099)
                /gene="STM1129"
                /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene       5599. .6759
                /gene="STM1130"
     CDS        5599. .6759
                /gene="STM1130"
                /note="similar to E. coli orf, hypothetical protein
                (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                identity in aa 40 - 403"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL20061.1"
                /db_xref="GI:16419646"
                /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene       6793. .7497
                /gene="STM1131"
     RBS        6793. .6798
                /gene="STM1131"
                /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
     CDS        6805. .7497
                /gene="STM1131"
                /note="similar to E. coli orf, hypothetical protein
                (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                identity in aa 4 - 239"
                /codon_start=1
                /transl_table=11
                /product="putative outer membrane protein"
                /protein_id="AAL20062.1"
                /db_xref="GI:16419647"
                /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                QTRYRVGLQYHF"
     gene       7768. .9060
                /gene="STM1132"
     RBS        7768. .7773
                /gene="STM1132"
                /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS        7780. .9060
                /gene="STM1132"
                /note="similar to E. coli putative transport protein
                (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
                identity in aa 25 - 418"
                /codon_start=1
                /transl_table=11
                /product="putative sugar transport protein"

  Query Match             56.8%;  Score 259.2;  DB 1;  Length 22411;
  Best Local Similarity   73.0%;  Pred. No. 3.6e-64;
  Matches  333;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db      17769 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 17828

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db      17829 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 17888

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  ||||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db      17889 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 17948

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              | |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db      17949 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 18008

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| ||||  | || || ||||| ||   ||| ||| | ||||
Db      18009 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 18068

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
```

```
Db      18069 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 18128

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |||| ||-   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db      18129 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 18188

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db      18189 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 18224


RESULT 15
AL627269
LOCUS       AL627269              254050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 5/20.
ACCESSION   AL627269 AL513382
VERSION     AL627269.1  GI:16502231
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            181. .2793
                     /gene="STY1078"
                     /note="synonym: pepN"
     CDS             181. .2793
                     /gene="STY1078"
                     /note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
                     to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aminopeptidase N"
                     /protein_id="CAD08183.1"
                     /db_xref="GI:16502232"
                     /db_xref="GOA:Q8Z7T0"
                     /db_xref="SPTREMBL:Q8Z7T0"
                     /translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
                     HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
                     NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
                     RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
                     NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
                     LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
                     GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
                     TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
                     KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
                     VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
                     AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
                     IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
                     ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
                     QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
                     AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
                     LSGDLYEKITKALA"
     misc_feature    532. .1332
                     /gene="STY1078"
                     /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
                     family M1, score 245.10, E-value 1.9e-84"
     misc_feature    1060. .1089
                     /gene="STY1078"
                     /note="PS00142 Neutral zinc metallopeptidases,
                     zinc-binding region signature"
     gene            3001. .4011
                     /gene="STY1079"
                     /note="synonym: pyrD"
     CDS             3001. .4011
                     /gene="STY1079"
                     /note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit
                     to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase"
                     /protein_id="CAD08184.1"
                     /db_xref="GI:16502233"
                     /db_xref="GOA:Q8Z7S9"
                     /db_xref="SWISS-PROT:Q8Z7S9"
                     /translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
                     KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
                     LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
                     KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
                     APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
                     EIIRRLSLELKGQLPIIGVGGIDSVIAARERIAAGATLVQIYSGFIFKGPPLIKEIVT
                     HI"
     misc_feature    3136. .3990
                     /gene="STY1079"
                     /note="Pfam match to entry PF01180 DHOdehase,
                     Dihydroorotate dehydrogenase, score 563.50, E-value
                     1.4e-165"
     misc_feature    3238. .3297
                     /gene="STY1079"
                     /note="PS00911 Dihydroorotate dehydrogenase signature 1"
     misc_feature    3874. .3936
                     /gene="STY1079"
                     /note="PS00912 Dihydroorotate dehydrogenase signature 2"
     gene            4174. .4719
                     /gene="STY1080"
     CDS             4174. .4719
                     /gene="STY1080"
                     /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                     YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                     discrepency in position of translational start site."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08185.1"
                     /db_xref="GI:16502234"
                     /db_xref="SPTREMBL:Q8Z7S8"
                     /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                     FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                     GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                     IMNDRLKPQVHCHSFSLEQAV"
     gene            complement(4716. .5825)
                     /gene="STY1081"
     CDS             complement(4716. .5825)
                     /gene="STY1081"
```

```
                     /note="Orthologue of E. coli P75863; Fasta hit to P75863
                     (369 aa), 86% identity in 367 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08186.1"
                     /db_xref="GI:16502235"
                     /db_xref="GOA:Q8Z7S7"
                     /db_xref="SPTREMBL:Q8Z7S7"
                     /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                     FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                     PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                     RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                     KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                     LDDSVTPEKHPDGSVTIDWQGQTPCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                     LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
misc_feature         complement(4737. .4955)
                     /gene="STY1081"
                     /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                     iron-sulfur cluster binding domains, score 55.90, E-value
                     8.5e-13"
gene                 5924. .8032
                     /gene="STY1082"
CDS                  5924. .8032
                     /gene="STY1082"
                     /note="Similar to Haemophilus influenzae hypothetical
                     protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                     scores: E(): 0, 60.3% id in 710 aa
                     Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                     (702 aa), 92% identity in 702 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD08187.1"
                     /db_xref="GI:16502236"
                     /db_xref="GOA:Q8Z7S6"
                     /db_xref="SPTREMBL:Q8Z7S6"
                     /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                     YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                     LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                     HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                     RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                     NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                     EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                     DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                     RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                     VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                     MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                     LSPDFARNRQIHNCWLIRAA"
misc_feature         5927. .7060
                     /gene="STY1082"
                     /note="Pfam match to entry PF01170 UPF0020,
                     Uncharacterized protein family UPF0020, score 781.30,
                     E-value 3.7e-231"
misc_feature         6506. .6541
                     /gene="STY1082"
                     /note="PS01261 Uncharacterized protein family UPF0020
                     signature"
gene                 8045. .9952
                     /gene="STY1083"
CDS                  8045. .9952
                     /gene="STY1083"
                     /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                     526 aa overlap
                     Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                     overlap

Query Match           56.5%;  Score 257.6;  DB 1;  Length 254050;
Best Local Similarity  72.8%;  Pred. No. 1.5e-63;
Matches  332;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db      88906 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 88965

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || | |||||||||||||| ||    || ||||| |||| |  |||
Db      88966 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGG 89025

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  ||||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db      89026 CCGGATTCCACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 89085

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db      89086 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 89145

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || ||||| | ||| |||| | || || ||||| ||   ||| ||| | ||||
Db      89146 GTAGGCCAGGGTGCGGATAACAGTACTATTGAACTGACTCAGAATGGTTTCAGAAACAAT 89205

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db      89206 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGTCAATACGGCGGT 89265

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db      89266 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 89325

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db      89326 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 89361

Search completed: March 18, 2004, 15:02:31
Job time : 2313.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	3	AAC64619	Aac64619 Escherich
2	456	100.0	456	9	ACF36153	Acf36153 E. coli C
3	437.2	95.9	646	2	AAQ62647	Aaq62647 Fibronect
4	355.4	77.9	369	2	AAQ62646	Aaq62646 FNB curli
5	254.4	55.8	456	2	AAQ87467	Aaq87467 AgfA sequ
6	254.4	55.8	456	2	AAT74142	Aat74142 Salmonell
7	254.4	55.8	456	3	AAC64617	Aac64617 Salmonell
8	238.4	52.3	456	3	AAC64628	Aac64628 AgfA::PT3
9	236.8	51.9	456	3	AAC64626	Aac64626 AgfA::PT3
10	230.4	50.5	456	3	AAC64625	Aac64625 AgfA::PT3
11	230.4	50.5	456	3	AAC64627	Aac64627 AgfA::PT3
12	228.8	50.2	456	3	AAC64630	Aac64630 AgfA::PT3
13	227.2	49.8	456	3	AAC64631	Aac64631 AgfA::PT3
14	225.6	49.5	456	3	AAC64623	Aac64623 AgfA::PT3
15	222.4	48.8	456	3	AAC64624	Aac64624 AgfA::PT3
16	220.8	48.4	456	3	AAC64629	Aac64629 AgfA::PT3
17	216	47.4	456	3	AAC64622	Aac64622 AgfA::PT3
18	173.4	38.0	361	2	AAQ73066	Aaq73066 AgfA sequ
19	173.4	38.0	361	2	AAT74141	Aat74141 Salmonell
20	100	21.9	100	7	ACD68807	Acd68807 E. coli K
21	100	21.9	100	7	ACD68808	Acd68808 E. coli K
22	100	21.9	100	7	ACD68809	Acd68809 E. coli K
23	78	17.1	78	9	ACF36151	Acf36151 E. coli C
24	72	15.8	72	9	ACF36148	Acf36148 E. coli C
25	66	14.5	66	9	ACF36150	Acf36150 E. coli C
26	43.2	9.5	1083	5	AAS76745	Aas76745 DNA encod
27	40.2	8.8	4018	2	AAQ63879	Aaq63879 Polyhydro
28	39	8.6	1390	9	ADC86984	Adc86984 Human GPC
c 29	39	8.6	2730	5	AAS87782	Aas87782 DNA encod
c 30	37.4	8.2	1972	9	ADC87176	Adc87176 Human GPC
31	37	8.1	2580	5	AAH66822	Aah66822 C glutami
c 32	37	8.1	349980	5	AAH68529	Aah68529 C glutami
33	36	7.9	36	9	ACF36152	Acf36152 E. coli V
34	36	7.9	36	9	ACF36149	Acf36149 E. coli C
c 35	35.8	7.9	3984	4	ABL27701	Abl27701 Drosophil
36	35.8	7.9	12775	4	ABL27700	Abl27700 Drosophil
37	34.8	7.6	1200	5	AAS92828	Aas92828 DNA encod
38	34.6	7.6	1633	9	ADC87022	Adc87022 Human GPC
c 39	34.6	7.6	3893	6	ABK52154	Abk52154 Haemophil
40	34.6	7.6	155074	6	ABN85735	Abn85735 Human gen
c 41	34.2	7.5	1543	5	AAS92206	Aas92206 DNA encod
42	34.2	7.5	2000	7	ADA71938	Ada71938 Rice gene
43	34.2	7.5	3532	9	ADC87116	Adc87116 Human GPC
44	34.2	7.5	38918	9	ADC87242	Adc87242 Human GPC
c 45	34	7.5	570	4	ABA36458	Aba36458 Probe #14

ALIGNMENTS

RESULT 1
AAC64619
ID AAC64619 standard; DNA; 456 BP.
XX
AC AAC64619;
XX
DT 26-FEB-2001 (first entry)
XX
DE Escherichia coli CsgA DNA sequence SEQ ID NO:3.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Escherichia coli.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36343.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

```
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 122 G; 111 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 456;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-134;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456


RESULT 2
ACF36153
ID   ACF36153 standard; DNA; 456 BP.
XX
AC   ACF36153;
XX
DT   04-DEC-2003  (first entry)
XX
DE   E. coli CsgA subunit 15 kDa protein encoding DNA.
XX
KW   Plasma protein; immune response; antibacterial; vaccine; gene therapy;
KW   gene; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .456
FT                   /*tag=  a
FT                   /note= "15 kDa protein"
XX
PN   WO2003064446-A2.
XX
PD   07-AUG-2003.
XX
PF   30-JAN-2003; 2003WO-EP000943.
XX
PR   31-JAN-2002; 2002GB-00002275.
XX
PA   (HANS-) HANSA MEDICAL RES AB.
XX
PI   Bjoerck L,  Olsen A,  Wikstroem M,  Herwald H;
XX
DR   WPI; 2003-646136/61.
DR   P-PSDB; ABR82651.
XX
PT   New isolated peptide capable of binding a mammalian plasma protein,
PT   useful in the manufacture of a medicament for the prevention and/or
PT   treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT   or Shigella infections.
XX
PS   Disclosure; Page 41-42; 42pp; English.
XX
CC   The invention relates to an isolated peptide capable of binding a
CC   mammalian plasma protein or of generating an immune response in a mammal
CC   selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC   antibody is useful for treating a bacterial infection in a human or
CC   animal or in the manufacture of a medicament for the prophylactic
CC   treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC   or Shigella infection. The peptide that is immobilized on a solid support
CC   is also useful as a reagent for determining the ability of a plasma
CC   protein to bind to bacteria. The present sequence represents an E. coli
CC   15 kDa protein encoding DNA
XX
SQ   Sequence 456 BP; 113 A; 110 C; 122 G; 111 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 456;  DB 9;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-134;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

```
RESULT 3
AAQ62647
ID   AAQ62647 standard; DNA; 646 BP.
XX
AC   AAQ62647;
XX
DT   25-MAR-2003  (revised)
DT   18-OCT-1994  (first entry)
XX
DE   Fibronectin binding protein fnbA gene.
XX
KW   Fibronectin binding protein; FBP; curlin; curli pili; pilus;
KW   immunization; vaccine; fnbA gene; ss.
XX
OS   Escherichia coli.
XX
PN   WO9410330-A1.
XX
PD   11-MAY-1994.
XX
PF   03-NOV-1993;   93WO-US010547.
XX
PR   03-NOV-1992;   92US-00970846.
XX
PA   (REAT/) REA T S.
XX
PI   Normark S,  Olsen A;
XX
DR   WPI; 1994-167484/20.
DR   P-PSDB; AAR52664.
XX
PT   New fibronectin binding protein curlin from E. coli - used for the
PT   treatment of infections caused by E. coli and for immunisation against E.
PT   coli infection.
XX
PS   Disclosure; Fig 2; 33pp; English.
XX
CC   The fnbA gene of E. coli cattle fecal isolate AO12 includes DNA encoding
CC   a new fibronectin binding protein (FBP), curlin, which is produced as
CC   curli pili when expressed in E. coli (pFnb20) DSM 4585. The fnbA gene
CC   sequence from pFnb20 is given in AAQ62647, and includes the sequence
CC   encoding the 17 kDa curlin subunit (amino acids 21-142 of AAR52664).
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 646 BP; 168 A; 146 C; 161 G; 171 T; 0 U; 0 Other;

  Query Match           95.9%;  Score 437.2;  DB 2;  Length 646;
  Best Local Similarity   98.9%;  Pred. No. 5.4e-128;
  Matches  451;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGAGCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTC--TCGTCAACGTGACTCAGGTT 500

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||
Db        501 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 536


RESULT 4
AAQ62646
ID   AAQ62646 standard; DNA; 369 BP.
XX
AC   AAQ62646;
XX
DT   25-MAR-2003  (revised)
DT   18-OCT-1994  (first entry)
XX
DE   FNB curlin.
XX
KW   Fibronectin binding protein; FBP; curlin; curli pili; pilus;
KW   immunization; vaccine; ss.
XX
OS   Escherichia coli.
XX
PN   WO9410330-A1.
XX
PD   11-MAY-1994.
XX
PF   03-NOV-1993;   93WO-US010547.
XX
PR   03-NOV-1992;   92US-00970846.
XX
PA   (REAT/) REA T S.
XX
PI   Normark S,  Olsen A;
XX
DR   WPI; 1994-167484/20.
DR   P-PSDB; AAR52663.
XX
PT   New fibronectin binding protein curlin from E. coli - used for the
PT   treatment of infections caused by E. coli and for immunisation against E.
PT   coli infection.
XX
PS   Claim 8; Page 22; 33pp; English.
XX
CC   The fnbA gene of E. coli cattle fecal isolate AO12 includes DNA encoding
CC   a new fibronectin binding protein, curlin, which is produced as curli
CC   pili when expressed in E. coli (pFnb20) DSM 4585. (Updated on 25-MAR-2003
CC   to correct PN field.)
XX
SQ   Sequence 369 BP; 86 A; 87 C; 105 G; 91 T; 0 U; 0 Other;

  Query Match           77.9%;  Score 355.4;  DB 2;  Length 369;
  Best Local Similarity   99.2%;  Pred. No. 3.5e-102;
  Matches  368;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy         60 AGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGG 119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGG 60

Qy        120 CCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCA 120

Qy        180 AACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGA 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGA 180

Qy        240 TGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAG 299
```

```
Db        181 TGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAG 240

Qy        300 CGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGG 359

Db        241 CGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGG 300

Qy        360 TGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGT 419

Db        301 TGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTC--TCGTCAACGTGACTCAGGT 358

Qy        420 TGGCTTTGGTA 430

Db        359 TGGCTTTGGTA 369
```

RESULT 5
AAQ87467
ID AAQ87467 standard; DNA; 456 BP.
XX
AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE AgfA sequence.
XX
KW Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS Salmonella.
XX
FH Key Location/Qualifiers
FT CDS 1. .454
FT /*tag= a
FT /note= "AgfA"
XX
PN WO9425598-A2.
XX
PD 10-NOV-1994.
XX
PF 26-APR-1994; 94WO-IB000207.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
DR WPI; 1994-358275/44.
DR P-PSDB; AAR74625.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7B; 95pp; English.
XX
CC The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC proteins are used in genetic immunization and vaccine compositions,
CC respectively, to elicit an immune response to Salmonella in animals (e.g.
CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 55.8%; Score 254.4; DB 2; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.9e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 6
AAT74142
ID AAT74142 standard; DNA; 456 BP.
XX
AC AAT74142;
XX
DT 25-MAR-2003 (revised)
DT 29-SEP-1997 (first entry)
XX
DE Salmonella enteritidis 27655-3b agfA gene.
XX
KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS Salmonella enteritidis.
XX
FH Key Location/Qualifiers
FT CDS 1. .456
FT /*tag= a
FT /label= agfA_gene_fragment
FT /transl_except= (pos:367. .369,aa:Pro)
XX
PN US5635617-A.
XX
PD 03-JUN-1997.
XX
PF 26-APR-1994; 94US-00233788.
XX
PR 26-APR-1993; 93US-00054452.
XX
PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI Collinson SK, Kay WW, Doran JL;
XX
DR WPI; 1997-309886/28.
DR P-PSDB; AAW23570.
XX
PT Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT enteropathogenic bacteria of the Enterobacteria family.
XX
PS Claim 1; Col 19-112; 85pp; English.
XX
CC The present sequence represents an isolated agfA gene derived from
CC Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC diagnostic assays for Salmonella and/or enteropathogenic bacteria of the

```
CC   family Enterobacteria. It can also be used to provide proteins and
CC   antibodies which can be used for assays. The nucleic acid sequence can be
CC   used to provide probes or primers which can specifically hybridise to
CC   nucleic acid molecules from greater than 99% of Salmonella strains that
CC   are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC   from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC   correct PF field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match           55.8%;  Score 254.4;  DB 2;  Length 456;
  Best Local Similarity  72.4%;  Pred. No. 3.9e-70;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || | |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 7
AAC64617
ID   AAC64617 standard; DNA; 456 BP.
XX
AC   AAC64617;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36341.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 134; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match           55.8%;  Score 254.4;  DB 3;  Length 456;
  Best Local Similarity  72.4%;  Pred. No. 3.9e-70;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || | |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
```

```
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 8
AAC64628
ID   AAC64628 standard; DNA; 456 BP.
XX
AC   AAC64628;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36352.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

  Query Match           52.3%;  Score 238.4;  DB 3;  Length 456;
  Best Local Similarity  70.2%;  Pred. No. 4.5e-65;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| | |||   | |||||  |   | | |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                  |    ||    |   | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 9
AAC64626
ID   AAC64626 standard; DNA; 456 BP.
XX
AC   AAC64626;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36350.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
```

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 51.9%; Score 236.8; DB 3; Length 456;
Best Local Similarity 70.0%; Pred. No. 1.5e-64;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| | |||  |  ||||||  ||  |  | |  |||  ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 10
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 50.5%; Score 230.4; DB 3; Length 456;
Best Local Similarity 69.1%; Pred. No. 1.6e-62;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | | ||| | |    ||  ||     | |  |  | |    |||| ||| ||||||
```

```
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 11
AAC64627
ID   AAC64627 standard; DNA; 456 BP.
XX
AC   AAC64627;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36351.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 137; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

  Query Match             50.5%;  Score 230.4;  DB 3;  Length 456;
  Best Local Similarity   69.1%;  Pred. No. 1.6e-62;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Db        241 GTAGGCCAGGGTGCGGATAATTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATG 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Db        301 GCACATGCAGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456

Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
AAC64630
ID   AAC64630 standard; DNA; 456 BP.
XX
AC   AAC64630;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
```

PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36354.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 50.2%; Score 228.8; DB 3; Length 456;
Best Local Similarity 68.9%; Pred. No. 5e-62;
Matches 314; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||   | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
                || ||||                | | ||   |    || ||  ||| | || |
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
```

```
                || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36355.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response

```
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

  Query Match             49.8%;  Score 227.2;  DB 3;  Length 456;
  Best Local Similarity   68.6%;  Pred. No. 1.6e-61;
  Matches  313;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||   |||||| |    |   | |          |  |||    || |||
Db        301 GCCACCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||  ||   ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 225.6;  DB 3;  Length 456;
  Best Local Similarity   68.4%;  Pred. No. 5.2e-61;
  Matches  312;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    |||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||   | || |||||||||  ||||| || || || || || ||    || | | |
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATATGATCAG 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                          ||  |||  |||         |  |  |   |||  |  |||  |||||||
Db        361 CTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCAAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 15
AAC64624
ID   AAC64624 standard; DNA; 456 BP.
XX
AC   AAC64624;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36348.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal..
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

  Query Match             48.8%;  Score 222.4;  DB 3;  Length 456;
  Best Local Similarity   68.0%;  Pred. No. 5.3e-60;
  Matches  310;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || ||  | |||| |  | | | |  | ||  |  ||  ||  ||||
Db         61 GGCGTCTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCATCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||  | ||| |||||| |||||||| || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||  ||| |||||||||  ||  |||||  || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||  ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| ||||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456

Search completed: March 18, 2004, 09:11:49
Job time : 290.405 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		254.4	55.8	456	1	US-08-233-788A-58	Sequence 58, Appl
2		173.4	38.0	361	1	US-08-233-788A-56	Sequence 56, Appl
3	c	35.6	7.8	1788	4	US-09-252-991A-4632	Sequence 4632, Ap
4		35.6	7.8	2250	4	US-09-252-991A-4181	Sequence 4181, Ap
5	c	35.6	7.8	3102	4	US-09-252-991A-4429	Sequence 4429, Ap
6	c	34	7.5	552	4	US-09-252-991A-2024	Sequence 2024, Ap
7		34	7.5	669	4	US-09-252-991A-1852	Sequence 1852, Ap
8		34	7.5	1185	4	US-09-252-991A-1931	Sequence 1931, Ap
9		33.4	7.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
10		33.4	7.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
11		33	7.2	1830121	4	US-09-557-884-1	Sequence 1, Appli
12		33	7.2	1830121	4	US-09-643-990A-1	Sequence 1, Appli
13	c	32	7.0	2993	2	US-08-225-488-6	Sequence 6, Appli
14	c	31.8	7.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
15	c	31.4	6.9	1317	4	US-09-252-991A-10989	Sequence 10989, A
16		31.2	6.8	1830	4	US-09-667-135-29	Sequence 29, Appl
17		31	6.8	1422	4	US-09-328-352-586	Sequence 586, App
18		30.6	6.7	435	3	US-08-569-147-75	Sequence 75, Appl
19		30.6	6.7	1923	4	US-09-056-556-190	Sequence 190, App
20		30.6	6.7	1923	4	US-09-072-596-185	Sequence 185, App
21		30.6	6.7	1923	4	US-09-072-967-190	Sequence 190, App
22	c	30.6	6.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
23	c	30	6.6	6122	1	US-08-403-545-1	Sequence 1, Appli
24	c	30	6.6	6122	3	US-08-404-381-1	Sequence 1, Appli
25	c	30	6.6	1230025	4	US-09-198-452A-1	Sequence 1, Appli
26		29.8	6.5	771	4	US-09-489-039A-915	Sequence 915, App
27		29.8	6.5	1128	4	US-09-328-352-3993	Sequence 3993, Ap
28		29.8	6.5	2652	4	US-09-489-039A-5584	Sequence 5584, Ap
29	c	29.8	6.5	87563	4	US-09-453-702B-57	Sequence 57, Appl
30		29.2	6.4	14654	4	US-08-961-527-106	Sequence 106, App
31	c	28.8	6.3	429	4	US-09-252-991A-8302	Sequence 8302, Ap
32	c	28.8	6.3	912	2	US-08-993-228-7	Sequence 7, Appli
33		28.8	6.3	1023	4	US-09-328-352-3184	Sequence 3184, Ap
34		28.8	6.3	1395	4	US-09-252-991A-8312	Sequence 8312, Ap
35		28.8	6.3	9507	4	US-09-489-039A-6896	Sequence 6896, Ap
36		28.8	6.3	12720	1	US-08-403-866-11	Sequence 11, Appl
37		28.6	6.3	217	4	US-09-313-294A-5247	Sequence 5247, Ap
38		28.6	6.3	693	4	US-09-328-352-3008	Sequence 3008, Ap
39		28.6	6.3	1507	4	US-09-453-323-1	Sequence 1, Appli
40		28.6	6.3	1873	3	US-09-461-474-9	Sequence 9, Appli
41		28.4	6.2	1380	4	US-09-489-039A-2442	Sequence 2442, Ap
42	c	28.4	6.2	62909	4	US-09-596-002-32	Sequence 32, Appl
43		28.2	6.2	298	4	US-09-313-294A-4713	Sequence 4713, Ap
44		28.2	6.2	346	3	US-09-060-756-317	Sequence 317, App
45		28.2	6.2	346	4	US-09-670-314-317	Sequence 317, App

ALIGNMENTS

RESULT 1
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-788A-58

Query Match 55.8%; Score 254.4; DB 1; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.3e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456
```

```
RESULT 2
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..357
US-08-233-788A-56

  Query Match             38.0%;  Score 173.4;  DB 1;  Length 361;
  Best Local Similarity   69.9%;  Pred. No. 7.3e-45;
  Matches  234;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         64 GTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCA 123
Db          1 GTCGTACCACAGTGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGCCCG 60

Qy        124 AATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACT 183
Db         61 GACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAAAGC 120

Qy        184 GATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTT 243
Db        121 GATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTA 180

Qy        244 GGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCT 303
Db        181 GGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCC 240

Qy        304 ACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGC 363
Db        241 ACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGTAAT 300

Qy        364 AACGGTGCTGCAGTTGACCAGACTGCATCTAACTC 398
Db        301 AACGCCGCGCTGGTTAATCAGACCGCATCTGATTC 335


RESULT 3
US-09-252-991A-4632/c
; Sequence 4632, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 4632
;   LENGTH: 1788
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4632

  Query Match             7.8%;  Score 35.6;  DB 4;  Length 1788;
  Best Local Similarity   64.6%;  Pred. No. 0.23;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         30 AGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCCGGCGG 89
Db       1391 AGCGGGTGTATTCGCCGGTGCCGGACTGCGCGGTGGCGATCGGCAGGAAGCCGGCGGCGG 1332

Qy         90 TAACCACGGTGGTGGCGGTAAT 111
Db       1331 TGATCAGGGTGCCGGTGAGCAT 1310
```

```
RESULT 4
US-09-252-991A-4181
; Sequence 4181, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 4181
;   LENGTH: 2250
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4181

  Query Match             7.8%;  Score 35.6;  DB 4;  Length 2250;
  Best Local Similarity   64.6%;  Pred. No. 0.26;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         30 AGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGG 89
              |||    |||||| |||||  ||  |||   |||| | ||  ||| | | |||||||||
Db        339 AGCGGGTGTATTCGCCGGTGCCGGACTGCGCGGTGGCGATCGGCAGGAAGCCGGCGGCGG 398

Qy         90 TAACCACGGTGGTGGCGGTAAT 111
              | | || ||||  || |   ||
Db        399 TGATCAGGGTGCCGGTGAGCAT 420


RESULT 5
US-09-252-991A-4429/c
; Sequence 4429, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 4429
;   LENGTH: 3102
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4429

  Query Match             7.8%;  Score 35.6;  DB 4;  Length 3102;
  Best Local Similarity   64.6%;  Pred. No. 0.3;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         30 AGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGG 89
              |||    |||||| |||||  ||  |||   |||| | ||  ||| | | |||||||||
Db       1411 AGCGGGTGTATTCGCCGGTGCCGGACTGCGCGGTGGCGATCGGCAGGAAGCCGGCGGCGG 1352

Qy         90 TAACCACGGTGGTGGCGGTAAT 111
              | | || ||||  || |   ||
Db       1351 TGATCAGGGTGCCGGTGAGCAT 1330


RESULT 6
US-09-252-991A-2024/c
; Sequence 2024, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2024
;   LENGTH: 552
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2024

  Query Match             7.5%;  Score 34;  DB 4;  Length 552;
  Best Local Similarity   56.1%;  Pred. No. 0.44;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        147 GTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGAC 206
              || ||| |||| |  |  ||  | || || ||   | ||||| ||   |  ||   |  |
Db        119 GTGCGGCGGCGCTGGCGGTGTTCGTGTTCGGCGTCCCGATGCGCGGCTCCTTGCTGTCCC 60

Qy        207 TATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGA 260
              | || | |      ||| | | | |||||||  |  | |||     || || | |||
Db         59 TGTTGCTCTTGTCGGCCGTGTTCATGGTGCCGGCGCTGGGTCAGGGTCTGCTGA 6


RESULT 7
US-09-252-991A-1852
; Sequence 1852, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1852
;   LENGTH: 669
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1852

  Query Match             7.5%;  Score 34;  DB 4;  Length 669;
  Best Local Similarity   56.1%;  Pred. No. 0.48;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        147 GTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGAC 206
              || ||| |||| |  |  ||  | || || ||   | ||||| ||   |  ||   |  |
Db        462 GTGCGGCGGCGCTGGCGGTGTTCGTGTTCGGCGTCCCGATGCGCGGCTCCTTGCTGTCCC 521

Qy        207 TATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGA 260
              | || | |      ||| | | | |||||||  |  | |||     || || | |||
Db        522 TGTTGCTCTTGTCGGCCGTGTTCATGGTGCCGGCGCTGGGTCAGGGTCTGCTGA 575


RESULT 8
US-09-252-991A-1931
; Sequence 1931, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
```

```
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1931
;    LENGTH: 1185
;    TYPE: DNA
;    ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1931

  Query Match                 7.5%;  Score 34;  DB 4;  Length 1185;
  Best Local Similarity   56.1%;  Pred. No. 0.62;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        147 GTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGAC 206
              || ||| |||| | |  |  || | ||  ||  | |||||  ||   |  ||   | |
Db        782 GTGCGGCGGCGCTGGCGGTGTTCGTGTTCGGCGTCCCGATGCGCGGCTCCTTGCTGTCCC 841

Qy        207 TATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGA 260
              | || | |  ||| | | | ||||||| | | |||   || | | | |||
Db        842 TGTTGCTCTTGTCGGCCGTGTTCATGGTGCCGGCGCTGGGTCAGGGTCTGCTGA 895


RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;    LENGTH: 4403765
;    TYPE: DNA
;    ORGANISM: Mycobacterium tuberculosis
;    FEATURE:
;    OTHER INFORMATION: CDC 1551
;    OTHER INFORMATION: "n" bases at various positions throughout the sequence
;    OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match                 7.3%;  Score 33.4;  DB 3;  Length 4403765;
  Best Local Similarity   72.9%;  Pred. No. 30;
  Matches   43;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        206 CTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGC 264
              || |  | ||||| | ||||| || |||||| |||||||| |||| || |  ||| ||
Db     371246 CTGTCGCGCAGCAGGCCGGCGATACTGGTGCTGATGTTGGCCAGGCCTGAATTGAAGGC 371304


RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;    LENGTH: 4411529
;    TYPE: DNA
;    ORGANISM: Mycobacterium tuberculosis
;    OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match                 7.3%;  Score 33.4;  DB 3;  Length 4411529;
  Best Local Similarity   72.9%;  Pred. No. 30;
  Matches   43;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        206 CTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGC 264
              || |  | ||||| | ||||| || |||||| |||||||| |||| || |  ||| ||
Db     371189 CTGTCGCGCAGCAGGCCGGCGATACTGGTGCTGATGTTGGCCAGGCCTGAATTGAAGGC 371247


RESULT 11
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;  GENERAL INFORMATION:
;       APPLICANT: Fleischmann et al.
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/557,884
;            FILING DATE: 25-Apr-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/476,102
;            FILING DATE: JUN-5-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Michelle S. Marks
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB186P3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-309-8504
;            TELEFAX: 301-309-8439
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match                 7.2%;  Score 33;  DB 4;  Length 1830121;
  Best Local Similarity   55.8%;  Pred. No. 30;
  Matches   63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
```

```
                 | | || |     || | ||||  | |    | |   || |||||  | | | |||| |
Db     740925 CGATTTTTCGCAAGATCGTTTAAATGAATATCGCCTACAATTCTGCCTTGGTTATGCACA 740984

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGT 234
              ||||| | ||   |   ||   | |||||    | ||| ||||| ||| |||
Db     740985 TTGATGTCTGTTTTTAAAACCAAATTATTATTTTGAATGTCGGCGTTAAGGGT 741037


RESULT 12
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;        APPLICANT: Robert D. Fleischmann
;                   Mark D. Adams
;                   Owen White
;                   Hamilton O. Smith
;                   J. Craig Venter
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville,
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/643,990A
;             FILING DATE: 23-Aug-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/487,429
;             FILING DATE: 1995-06-07
;             APPLICATION NUMBER: 08/426,787
;             FILING DATE: 1995-04-21
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB186P1C1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-610-5790
;             TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match             7.2%;  Score 33;  DB 4;  Length 1830121;
  Best Local Similarity   55.8%;  Pred. No. 30;
  Matches   63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
              | | || |     || | ||||  | |    | |   || |||||  | | | |||| |
Db     740925 CGATTTTTCGCAAGATCGTTTAAATGAATATCGCCTACAATTCTGCCTTGGTTATGCACA 740984

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGT 234
              ||||| | ||   |   ||   | |||||    | ||| ||||| ||| |||
Db     740985 TTGATGTCTGTTTTTAAAACCAAATTATTATTTTGAATGTCGGCGTTAAGGGT 741037


RESULT 13
US-08-225-488-6/c
; Sequence 6, Application US/08225488
; Patent No. 5846802
;  GENERAL INFORMATION:
;    APPLICANT:  Buxton, Frank
;    APPLICANT:  Hinnen, Albert
;    APPLICANT:  Visser, Jacob
;    TITLE OF INVENTION:  No. 5846802el Fungal Protease
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/225,488
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/047,214
;      FILING DATE:  13-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  4-19055/A/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2993 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;      STRAIN:  N400
;    IMMEDIATE SOURCE:
;      CLONE:  pTZPEPD
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..829
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(830..1153, 1205..1649, 1697..1785, 1841
;      LOCATION:  ..2233)
;      OTHER INFORMATION:  /note= "subtilisin-type protease
;      OTHER INFORMATION:  PEPD of Aspergillus niger; product of gene pepD"
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1154..1204
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1650..1696
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1786..1840
US-08-225-488-6

  Query Match             7.0%;  Score 32;  DB 2;  Length 2993;
  Best Local Similarity   46.4%;  Pred. No. 3.9;
  Matches  104;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;
```

```
Qy         65 TTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAA 124
              ||||    ||| |    || |        ||||  ||||| | | | |     |  ||| |
Db       2181 TTGTAGGCCAGAAGATTGGGGCTACCCGCCACATTGGTGACAGCATTCCGCGTAGCCAAC 2122

Qy        125 ATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTG 184
               |  ||||||      || |    | | ||  | ||      | |  || |  |||| |
Db       2121 CTCTTGAGCTCGGTCGTTGCAGCCGCTGGGGTAGCAAGGTCCCGCAAGCCCATCAAATAG 2062

Qy        185 ATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTG 244
              | |  |  | || | || |||    || |||     | | |||  || |||  | ||
Db       2061 AGGATCAAACCTGTCACATGAGGTGTAGCCATGGACGTGCCGGAGATCGTGTTGGTGGCC 2002

Qy        245 GTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGC 288
              |     |  | | |    ||  ||   |  ||  |   || |||
Db       2001 GAGTTCGAGCCGGTCCATGCAGAAAGTACTTGCTCTCCCGGGGC 1958


RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

 Query Match             7.0%;  Score 31.8;  DB 3;  Length 4403765;
  Best Local Similarity   46.0%;  Pred. No. 72;
  Matches  108;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy         82 GGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATT 141
              ||||  ||     ||| |||||  |||   |||  ||||         |||       |
Db    4029096 GGCGTTGGGGGTGACGCTGGTGACGGGGGTAACGGCGGGATCGGCGGTGACGGCGGCGCC 4029037

Qy        142 TACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGAC 201
               |    |||| ||||    |   ||    |  |||       |   || ||| ||| |
Db    4029036 GGCGGAGACGGCGGCGCGGGCGGCGCGGGCGGCCTGTTCGGCGTCTCCGGTAGCTCCGGG 4028977

Qy        202 TTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGAC 261
              ||      | ||      | |||| |||||||| |  |  | |||| ||  |   |  | |
Db    4028976 CTGGGCGGTGCCGCCGGTAGCGGGGGTAATGGCGGGGGCGGTGGTGAGCCCGGTGTGGCC 4028917

Qy        262 AGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGT 316
              ||||    |  |  |    |||  ||  |  ||  |  || |  ||||
Db    4028916 GGCTCCCCGGGCGTTGGGCCTGCGGGCCGTGGCGGCGACGGCAACCTCGGCCAGT 4028862


RESULT 15
US-09-252-991A-10989/c
; Sequence 10989, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10989
;   LENGTH: 1317
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10989

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 1317;
  Best Local Similarity   52.7%;  Pred. No. 4.2;
  Matches   68;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         64 GTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCA 123
              ||||| | |||| | ||| |     |||  |    || || | |||  |         |||
Db        513 GTTGTCCTTCAGCAGGGCCGGCAGGGTGTCGTGGGAGGGGACGGCGACGTCGATCGCCTC 454

Qy        124 AATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACT 183
                 ||| || |  | ||  |||||  |    ||||||   |    |   | |  ||  | |
Db        453 GCCACTGCGCAGCGCCTTGTCCAGTTCCTCGGCGGTGGCGAAGGTGTGGTACTCGACGCG 394

Qy        184 GATGCCCGT 192
              |||||| ||
Db        393 GATGCCGGT 385


Search completed: March 18, 2004, 20:20:03
Job time : 76.4971 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	42.6	9.3	135259	12	US-10-240-425-1585	Sequence 1585, Ap
	2	39	8.6	1390	14	US-10-017-161-1781	Sequence 1781, Ap
	3	39	8.6	1390	15	US-10-292-798-1437	Sequence 1437, Ap
	4	38.2	8.4	625	14	US-10-029-386-24675	Sequence 24675, A
c	5	37.4	8.2	1972	14	US-10-017-161-1981	Sequence 1981, Ap
c	6	37.4	8.2	1972	15	US-10-292-798-1629	Sequence 1629, Ap
	7	37	8.1	2580	9	US-09-738-626-1857	Sequence 1857, Ap
c	8	37	8.1	3309400	9	US-09-738-626-1	Sequence 1, Appli
c	9	35.8	7.9	1493	14	US-10-029-386-25133	Sequence 25133, A
c	10	35.6	7.8	992	12	US-10-425-114-31747	Sequence 31747, A
	11	35.6	7.8	2122	15	US-10-108-260A-1242	Sequence 1242, Ap
c	12	35.6	7.8	3078	12	US-10-389-647-226	Sequence 226, App
	13	35.6	7.8	4389	15	US-10-369-493-46798	Sequence 46798, A
c	14	35	7.7	785	14	US-10-029-386-22627	Sequence 22627, A
	15	34.6	7.6	532	14	US-10-029-386-7202	Sequence 7202, Ap
	16	34.6	7.6	1633	14	US-10-017-161-1819	Sequence 1819, Ap
	17	34.6	7.6	1633	15	US-10-292-798-1475	Sequence 1475, Ap
	18	34.6	7.6	155074	13	US-10-026-188-6	Sequence 6, Appli
c	19	34.2	7.5	528	14	US-10-029-386-5350	Sequence 5350, Ap
c	20	34.2	7.5	666	14	US-10-029-386-25948	Sequence 25948, A
	21	34.2	7.5	1692	14	US-10-355-430-33	Sequence 33, Appl
	22	34.2	7.5	3532	14	US-10-017-161-1913	Sequence 1913, Ap
	23	34.2	7.5	3532	15	US-10-292-798-1569	Sequence 1569, Ap
c	24	34.2	7.5	12981	15	US-10-369-493-33865	Sequence 33865, A
	25	34.2	7.5	38918	14	US-10-017-161-2049	Sequence 2049, Ap
	26	34.2	7.5	38918	15	US-10-292-798-1695	Sequence 1695, Ap
c	27	34	7.5	570	9	US-09-864-761-21778	Sequence 21778, A
c	28	34	7.5	1976	9	US-09-864-761-5052	Sequence 5052, Ap
	29	33.8	7.4	765	15	US-10-369-493-44503	Sequence 44503, A
c	30	33.6	7.4	652	15	US-10-260-238-4681	Sequence 4681, Ap
c	31	33.6	7.4	1197	14	US-10-156-761-4579	Sequence 4579, Ap
c	32	33.6	7.4	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	33	33.4	7.3	5163	12	US-10-282-122A-26364	Sequence 26364, A
c	34	33.4	7.3	6615	12	US-10-282-122A-28180	Sequence 28180, A
c	35	33.2	7.3	832	12	US-10-282-122A-19206	Sequence 19206, A
	36	33.2	7.3	1377	14	US-10-238-075-1261	Sequence 1261, Ap
c	37	33.2	7.3	11475	14	US-10-238-075-1250	Sequence 1250, Ap
c	38	33	7.2	379	14	US-10-029-386-19106	Sequence 19106, A
c	39	33	7.2	569	14	US-10-029-386-3710	Sequence 3710, Ap
c	40	33	7.2	652	14	US-10-029-386-22783	Sequence 22783, A
	41	33	7.2	867	12	US-10-282-122A-19705	Sequence 19705, A
	42	33	7.2	1959	9	US-09-864-761-4012	Sequence 4012, Ap
c	43	33	7.2	1973	9	US-09-864-761-3471	Sequence 3471, Ap
c	44	33	7.2	3897	12	US-10-282-122A-22081	Sequence 22081, A
c	45	33	7.2	9590	12	US-10-282-122A-33351	Sequence 33351, A

ALIGNMENTS

RESULT 1
US-10-240-425-1585/c
; Sequence 1585, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1585
; LENGTH: 135259
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 Z83838
US-10-240-425-1585

Query Match 9.3%; Score 42.6; DB 12; Length 135259;
Best Local Similarity 47.5%; Pred. No. 0.011;
Matches 160; Conservative 0; Mismatches 174; Indels 3; Gaps 1;

```
Qy          47 GTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCG 106
               ||| | ||  | | || || | |   | | || || | | | ||| ||| ||| |
Db       70242 GTAATGGTGGTGGTGATGGTGATGGTGGTAATGATGGTGGTGCTGATCACAGTGATGGTG 70183
```

```
Qy        107 GTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTG 166
Db      70182 GTGGTGATGGTGATCATGGTGGTGATGGTGATGGTTGTAATGATGATGGTGATGATGGTG 70123

Qy        167 CACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCG 226
Db      70122 ATGGTGCAGACCACAGTGATG---GTGATGGTGGTGATGATAATGGTGATTATGGTGGTG 70066

Qy        227 GTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTG 286
Db      70065 ATTATGGTGATGGTGGTGTTGATGGTGCTGATCACAGTGATGGTGATAATGGCGATCATG 70006

Qy        287 GCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTA 346
Db      70005 GTTGTGGTGATGGTGGTGATGATGGTGCTGATCACAGTGATGGTGGTGGTGGTGACGATG 69946

Qy        347 AACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGACCA 383
Db      69945 GTGATCATGGTGGTGGTGATGGTGGTGATGGTGCTCA 69909
```

RESULT 2
US-10-017-161-1781
; Sequence 1781, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1781
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1390)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1190)
US-10-017-161-1781

Query Match 8.6%; Score 39; DB 14; Length 1390;
Best Local Similarity 45.1%; Pred. No. 0.021;
Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```
Qy         62 GTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCC 121
Db        853 GTGATGGTGGTGATGGTGATGGCAGTGGTGGCAGTGGTGGTGGTGGTGATAATGATGGTC 912

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
Db        913 ATGGTAGTGGTGGTGATGGTGATAGTGATGGTGATGGCAGTGGTGGCAGTGGTGGTGATA 972

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATG 241
Db        973 GTGATGGTCATGGCAGTGGTGGTGATGGTGATAGTGATGGTGATGGCAGTGGTGGCAGTG 1032

Qy        242 TTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCG 301
Db       1033 GTGGTGATGGTAATAGTGATGGTGATGGCAGTGATAGTGATAGTGATGGTGATGGCAGTG 1092

Qy        302 CTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTG 361
Db       1093 GTGGTGGTGGTGATAGTGATGGTGATGGCAGTGGTGGTGGTGGTGATAGTGATGGTGATG 1152
```

```
Qy        362 GCAACGGTGCTGCAGTTGA 380
Db       1153 GCAATGGTGGTGATGGTGA 1171
```

RESULT 3
US-10-292-798-1437
; Sequence 1437, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1437
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1390)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1190)
US-10-292-798-1437

Query Match 8.6%; Score 39; DB 15; Length 1390;
Best Local Similarity 45.1%; Pred. No. 0.021;
Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```
Qy         62 GTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCC 121
Db        853 GTGATGGTGGTGATGGTGATGGCAGTGGTGGCAGTGGTGGTGGTGGTGATAATGATGGTC 912

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
Db        913 ATGGTAGTGGTGGTGATGGTGATAGTGATGGTGATGGCAGTGGTGGCAGTGGTGGTGATA 972

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATG 241
Db        973 GTGATGGTCATGGCAGTGGTGGTGATGGTGATAGTGATGGTGATGGCAGTGGTGGCAGTG 1032

Qy        242 TTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCG 301
Db       1033 GTGGTGATGGTAATAGTGATGGTGATGGCAGTGATAGTGATAGTGATGGTGATGGCAGTG 1092

Qy        302 CTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTG 361
Db       1093 GTGGTGGTGGTGATAGTGATGGTGATGGCAGTGGTGGTGGTGGTGATAGTGATGGTGATG 1152

Qy        362 GCAACGGTGCTGCAGTTGA 380
Db       1153 GCAATGGTGGTGATGGTGA 1171
```

RESULT 4
US-10-029-386-24675
; Sequence 24675, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.

```
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24675
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP001754.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
;   OTHER INFORMATION: EST_HUMAN HIT: BI015471.1, EVALUE 1.00e-03
;   OTHER INFORMATION: NT HIT: AE000415.1, EVALUE 9.30e+00
US-10-029-386-24675

  Query Match             8.4%;  Score 38.2;  DB 14;  Length 625;
  Best Local Similarity   46.2%;  Pred. No. 0.026;
  Matches  162;  Conservative    0;  Mismatches  188;  Indels    1;  Gaps    1;

Qy         29 CAGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCG 88
              || ||  | || |     |||| |      | |||| || |  | |  | || ||| |||
Db         87 CATCACCCTTACTTATGAGTAGTGGGGGCGGTGGTGGTGATGGTGATGATGGTGGCTGCG 146

Qy         89 GTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGT 148
                |    | |||||||| ||| ||  ||| ||    | |  |||  |       | |   ||
Db        147 ATGGTGATGGTGGTGGTGGTGATGGTAGTGGTGGTAGTGGTGATGTTGGTGATGA-TGGT 205

Qy        149 ACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTA 208
                |||| ||| | | ||    || |       |||||   |||    ||       |
Db        206 GATGTGGTGGTGATTGTGATGGTGGTGGTGTTGGTGATGGTGGTAGTGATGGTGATGATG 265

Qy        209 TTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAA 268
                |    |   ||  ||| |||||||    | || || |   ||    ||||  ||
Db        266 ATGGTGATTCTGATGGCTGTAATGGAAGTGGTGGTGATGGTGGTGGTGATGGTAGTGGTG 325

Qy        269 TCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAA 328
                |||  |         ||    | ||    | | |  |  || | |  | |  ||
Db        326 GTGATGTTGATGGTGATGATGGCAGTGGTGGTGGTGGTGGTGGTGATGGTAGTGGTGGTG 385

Qy        329 ATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTG 379
              || |||         |||          ||||||||  | ||||  ||  | ||
Db        386 ATGTTGATGGTGATGTTGGCAGTGGTGGTGGTGGTGATGGTGGTGATGGTG 436


RESULT 5
US-10-017-161-1981/c
; Sequence 1981, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1981
;   LENGTH: 1972
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1972)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1772)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (975)..(994)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (1319)..(1328)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1981

  Query Match             8.2%;  Score 37.4;  DB 14;  Length 1972;
  Best Local Similarity   49.2%;  Pred. No. 0.086;
  Matches   98;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         62 GTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCC 121
              ||| |  |  | |  | || |  || |||  |   | ||||||  | |  | || | ||
Db        460 GTGATTATGATGATGATGGTGATGGTGGTGGCGGTGATGGTGGTGATGGTGATGGTGGTG 401

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
                  | ||            | |  |  | ||||| |||  || ||    || |    |
Db        400 ATGGTGGTGGTGGTGGTGGTGATGATGATGGTGGTGGTGGCTGTGATTATGATGATGATG 341

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATG 241
              |||||   ||  |  |||     | |    |  ||| || |||  |||||  ||||
Db        340 GTGATGGTGGTGGCGGTGATGGTGGTGATGGTGATGGTGGTGGTGGTGGTGGTGATGATG 281

Qy        242 TTGGTCAGGGCTCAGATGA 260
              ||||   ||||  ||| |
Db        280 GTGGTGGTGGCTGTGATTA 262


RESULT 6
US-10-292-798-1629/c
; Sequence 1629, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1629
;   LENGTH: 1972
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1972)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1772)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (975)..(994)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
```

```
;   LOCATION: (1319)..(1328)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-1629

  Query Match             8.2%;  Score 37.4;  DB 15;  Length 1972;
  Best Local Similarity   49.2%;  Pred. No. 0.086;
  Matches   98;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         62 GTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCC 121
Db        460 GTGATTATGATGATGATGGTGATGGTGGTGGCGGTGATGGTGGTGATGGTGATGGTGGTG 401

Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
Db        400 ATGGTGGTGGTGGTGGTGGTGATGATGATGGTGGTGGTGGCTGTGATTATGATGATGATG 341

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATG 241
Db        340 GTGATGGTGGTGGCGGTGATGGTGGTGATGGTGATGGTGGTGGTGGTGGTGATGATG 281

Qy        242 TTGGTCAGGGCTCAGATGA 260
Db        280 GTGGTGGTGGCTGTGATTA 262


RESULT 7
US-09-738-626-1857
; Sequence 1857, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1857
;   LENGTH: 2580
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1857

  Query Match             8.1%;  Score 37;  DB 9;  Length 2580;
  Best Local Similarity   47.8%;  Pred. No. 0.13;
  Matches  140;  Conservative    0;  Mismatches  150;  Indels    3;  Gaps    1;

Qy         77 ACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGA 136
Db       1046 AGGGTGACGCTGGCGCCGATGGCGAAGACGGCGCAGATGGTGAATCAATTACTGTTGTTA 1105

Qy        137 ACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTG---ATGCCCGTA 193
Db       1106 ACACCAGCAACGATGCTGAAGGCAACACTCTTGTTGAGCTCTCCGATGGCACTGTTATTA 1165

Qy        194 ACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCT 253
Db       1166 CCATTAATAAGGGTGATAAGGGTGACGCTGGCGCCGATGGTGAAGATGGAAGTAATGGCG 1225

Qy        254 CAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATC 313
Db       1226 CAGATGGCGAGTCCATTACGGTCATCGAAACCCGTTTCGATGCGGACGGCAACACTGTTG 1285

Qy        314 AGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAAC 366
Db       1286 TGGTGTTCTCGAACGGCACTGAAATCACCATCAACAAGGGTGAGAAGGGCGAC 1338


RESULT 8
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
;   LENGTH: 3309400
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

  Query Match             8.1%;  Score 37;  DB 9;  Length 3309400;
  Best Local Similarity   47.8%;  Pred. No. 4.5;
  Matches  140;  Conservative    0;  Mismatches  150;  Indels    3;  Gaps    1;

Qy         77 ACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGA 136
Db    1784428 AGGGTGACGCTGGCGCCGATGGCGAAGACGGCGCAGATGGTGAATCAATTACTGTTGTTA 1784369

Qy        137 ACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTG---ATGCCCGTA 193
Db    1784368 ACACCAGCAACGATGCTGAAGGCAACACTCTTGTTGAGCTCTCCGATGGCACTGTTATTA 1784309

Qy        194 ACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCT 253
Db    1784308 CCATTAATAAGGGTGATAAGGGTGACGCTGGCGCCGATGGTGAAGATGGAAGTAATGGCG 1784249

Qy        254 CAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATC 313
Db    1784248 CAGATGGCGAGTCCATTACGGTCATCGAAACCCGTTTCGATGCGGACGGCAACACTGTTG 1784189

Qy        314 AGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAAC 366
Db    1784188 TGGTGTTCTCGAACGGCACTGAAATCACCATCAACAAGGGTGAGAAGGGCGAC 1784136


RESULT 9
US-10-029-386-25133/c
; Sequence 25133, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25133
;   LENGTH: 1493
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP001751.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.6
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
US-10-029-386-25133

  Query Match             7.9%;  Score 35.8;  DB 14;  Length 1493;
  Best Local Similarity   44.2%;  Pred. No. 0.26;
  Matches  148;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy          46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
               |||   | |   |  |||| || |  | |    || || || ||| |    ||||||||
Db         987 GGTGATGGTGATGGTGGTGGTGATGGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGT 928

Qy         106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
               |||  | || |  |     |  ||       |   |   |||| | |     |
Db         927 GGTGGTGATTGTCGTGGTGGTGTTGGTGGTGGTGATGGTGATGGTGGTGGTGATGGTGAT 868

Qy         166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
               |  | || |    |   || ||  || |  || |     |  |       |||| ||
Db         867 GGTCATGATGGTGATGGTGGTGATGGTGATGGTGGCGGTGTTGGTGGTGGTGATGGTGGT 808

Qy         226 GGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGT 285
               ||| ||||||  | || ||||   ||    | ||   |        | |           ||
Db         807 GGTGATGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGATGGTGGTGATGGTGGT 748

Qy         286 GGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTT 345
               ||  | |||    | | |  | ||||   |  | ||  |   |  |||      | |
Db         747 GGTGTTGGTGGTGGTGATGGTGGTGATGGTGGTGATGGTGATGGTGGTGATGCTGATGGT 688

Qy         346 AAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGA 380
                   |   |||||||   | |||| ||  | |||
Db         687 GGTGATAGTGGTGGTGATGATGGTGGTGATGGTGA 653


RESULT 10
US-10-425-114-31747/c
; Sequence 31747, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 31747
;   LENGTH: 992
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMFLB73212F11_FLI
US-10-425-114-31747

  Query Match             7.8%;  Score 35.6;  DB 12;  Length 992;
  Best Local Similarity   54.6%;  Pred. No. 0.25;
  Matches   71;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          45 CGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGG 104
               |||   | ||| |  ||  ||| | ||    ||||||||||||      |||||| ||
Db         471 CGGGTCAGGTCTCGGCGGCCTTGGTGGTCTCGGCGGCGGCGGCGGCCTTGGCGGTGGCGG 412

Qy         105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
               |||   | | ||||| ||   || |  |||  |      |      | |||||||  |
Db         411 CGGCCTTGGTGGCGGCTCAGGATCGGGTCTGGGCGGCAGCGGTGGCCTTGGCGGTGGCAG 352

Qy         165 TGCACTTGCT 174
               ||  || | |
Db         351 TGGGCTCGGT 342


RESULT 11
US-10-108-260A-1242
; Sequence 1242, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1242
;   LENGTH: 2122
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-1242

  Query Match             7.8%;  Score 35.6;  DB 15;  Length 2122;
  Best Local Similarity   46.7%;  Pred. No. 0.37;
  Matches  113;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy          90 TAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTA 149
               || ||| | |       | || |||| |   |  |||||||  | ||||  || |||||
Db         302 TATCCAAGATCTCAAGGCTAGTAATTGGTGAGAAAATTCTGCACAGAACCATTCCCAGTG 361

Qy         150 CGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTAT 209
               | | | ||   | | |     || | | |||    | || |   | ||   || | |
Db         362 GTGAGACTGTTTTTGTCCTTCAGCCCGGGAAAGGCAGACCTGGCTCCCTCCTTTTATTTG 421

Qy         210 TACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAAT 269
               |  |  ||  |      |  |  |          |||||  | ||     ||| |
Db         422 CAAGCCTCAAAACCATCAAGACAGGCAAAGACACTCGAAGGGCGCGGACCTTCGCTGAGA 481

Qy         270 CGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAA 329
               ||       | | ||||     | |   |||| |||  ||| | || | | ||  || |
Db         482 CGCCACCCACGCAAGTGGAGGAAGCACTTGCGCCACTGCTGAGCGGTTGGAAGGGGAACA 541

Qy         330 TT 331
               |
Db         542 CT 543


RESULT 12
US-10-389-647-226/c
; Sequence 226, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
;  APPLICANT: GREENBERG, E. Peter
;  APPLICANT:  SCHUSTER, Martin
;  APPLICANT:  LOSTROH, Candi
;  TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
;  FILE REFERENCE: UIZ-038CP
;  CURRENT APPLICATION NUMBER: US/10/389,647
```

```
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: 09/653730
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: 60/153022
;  PRIOR FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 710
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 226
;   LENGTH: 3078
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-10-389-647-226

  Query Match             7.8%;  Score 35.6;  DB 12;  Length 3078;
  Best Local Similarity   64.6%;  Pred. No. 0.44;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy         30 AGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGG 89
              |||    |||||| ||||| ||  |||   |||| | ||  ||| | | ||||||||||
Db       1387 AGCGGGTGTATTCGCCGGTGCCGGACTGCGCGGTGGCGATCGGCAGGAAGCCGGCGGCGG 1328

Qy         90 TAACCACGGTGGTGGCGGTAAT 111
              | | || ||||  || |   ||
Db       1327 TGATCAGGGTGCCGGTGAGCAT 1306


RESULT 13
US-10-369-493-46798
; Sequence 46798, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46798
;   LENGTH: 4389
;   TYPE: DNA
;   ORGANISM: Bacillus subtilis
US-10-369-493-46798

  Query Match             7.8%;  Score 35.6;  DB 15;  Length 4389;
  Best Local Similarity   50.6%;  Pred. No. 0.52;
  Matches   86;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        247 CAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACT 306
              || |  |  ||  |  || | |  |  ||  |  |  |     ||| |||||||| ||
Db       2221 CATGAATTTGACAAAGGCCCGACGGTGCTTTCTGATTTTCTAAGCGGAAACAGCGCAACC 2280

Qy        307 CTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAAC 366
              |||||| |  |||| || |  | | ||||  | |||     | | | |   |||  |||
Db       2281 GTTGATCCGGCGAACCGCTATCACTTTGAAGCGCCGGAATTTCCGATTGTCAGTGCAAAC 2340

Qy        367 GGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCA 416
              |  | ||      |  || |       || || |  |||||   | | ||
Db       2341 GTAGATGTGTCAAATGAGCCCAAGCTGAAATCATTTGTCAAAAAGCCGCA 2390


RESULT 14
US-10-029-386-22627/c
; Sequence 22627, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22627
;   LENGTH: 785
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC007540.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.94
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.73
;   OTHER INFORMATION: EST_HUMAN HIT: BE767903.1, EVALUE 1.00e-03
;   OTHER INFORMATION: NT HIT: AJ252124.1, EVALUE 1.20e-02
US-10-029-386-22627

  Query Match             7.7%;  Score 35;  DB 14;  Length 785;
  Best Local Similarity   44.8%;  Pred. No. 0.36;
  Matches  134;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy         79 GGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAAC 138
              || |  || | | |  | ||||  || | | || ||   ||     ||  |||     |
Db       444 GGTGATGGTGATGATGATGGTGACGGTGATGATTATGATGGAGATGATGGTGATGGCGAT 385

Qy        139 ATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCT 198
              | | ||    | | |||  || |   ||     || |     |   || ||   || |   |
Db       384 AATGACGGAGATGATGGTAGTGATGATGGTGATGGTGATGATGATGGTGATGGTGATGAT 325

Qy        199 GACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGAT 258
              ||          |   |  ||  || | ||||| || ||||| || |   ||     | |
Db       324 GATGATGGAGATGATGATGGTGATGGTGATAATGATGGAGATGATGATGGTGGTCATGTT 265

Qy        259 GACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGTGG 318
              ||          |||       |  |||   ||| |   | | |  |||  | |
Db       264 GATGATGATGGTGATGGTGATGATGATGGTGATGGTGATGATGATGATGGAGATGATGGT 205

Qy        319 AACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTGCTGCAGT 377
              | ||| | |||  || |       | | |  |    |||| |||  | | ||   ||||
Db       204 GATGGCGATAATGATGGAGATGATGGTGATGATGATGGTGATGGTGATGATGTGACAGT 146


RESULT 15
US-10-029-386-7202
; Sequence 7202, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7202
;   LENGTH: 532
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC003693.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
```

; OTHER INFORMATION: NT HIT: AF143426.1, EVALUE 5.10e-01
; OTHER INFORMATION: SWISSPROT HIT: P31638, EVALUE 7.50e-01
US-10-029-386-7202

Query Match 7.6%; Score 34.6; DB 14; Length 532;
Best Local Similarity 48.3%; Pred. No. 0.41;
Matches 97; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

```
Qy          61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
               |||| || |  | |  | |     |  ||| |  | |||||||| |  ||| ||   ||
Db         294 GGTGGTGGTGGTGATGATGATATTGATGGTCATGATGGTGGTGGTGACAATGATCATGGG 353

Qy         121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
                 | ||||        |||   |  | ||||||||||     ||    || |  |||
Db         354 GACCGTGCTGATGGCGGTGTTTGTGATGATGGTGGCGGTGGTGGTGATGGTGGTGGGCAT 413

Qy         181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
               | |    |      |||    | ||       |  | || || | | || |||  | |
Db         414 GAGGGTAGTTGGGGTGGTGATGGCATTACCGTTGATGAGGGTGGTGATGATAGTGATGGT 473

Qy         241 GTTGGTCAGGGCTCAGATGAC 261
               | ||||   ||    ||||||
Db         474 GATGGTGGTGGGAATGATGAC 494
```

Search completed: March 19, 2004, 03:48:48
Job time : 263.734 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-3	Sequence 3, Appli
2	453	99.3	453	92	US-60-352-946-1	Sequence 1, Appli
3	453	99.3	453	101	US-60-444-371-1	Sequence 1, Appli
4	452.8	99.3	675	13	US-08-978-878-3	Sequence 3, Appli
5	452.8	99.3	675	31	US-09-741-873B-3	Sequence 3, Appli
6	396	86.8	396	13	US-08-978-878-1	Sequence 1, Appli
7	360	78.9	360	31	US-09-741-873B-1	Sequence 1, Appli
8	276.8	60.7	477	17	US-09-252-691-172	Sequence 172, App
9	276.8	60.7	477	17	US-09-252-691C-172	Sequence 172, App
10	276.8	60.7	477	51	US-10-417-886-172	Sequence 172, App
11	254.4	55.8	456	6	US-08-233-642A-56	Sequence 56, Appl
12	254.4	55.8	456	23	US-09-543-407-1	Sequence 1, Appli
13	238.4	52.3	456	23	US-09-543-407-23	Sequence 23, Appl
14	236.8	51.9	456	23	US-09-543-407-19	Sequence 19, Appl
15	230.4	50.5	456	23	US-09-543-407-17	Sequence 17, Appl
16	230.4	50.5	456	23	US-09-543-407-21	Sequence 21, Appl
17	228.8	50.2	456	23	US-09-543-407-27	Sequence 27, Appl
18	227.2	49.8	456	23	US-09-543-407-29	Sequence 29, Appl
19	225.6	49.5	456	23	US-09-543-407-13	Sequence 13, Appl
20	222.4	48.8	456	23	US-09-543-407-15	Sequence 15, Appl
21	220.8	48.4	456	23	US-09-543-407-25	Sequence 25, Appl
22	216	47.4	456	23	US-09-543-407-11	Sequence 11, Appl
23	173.4	38.0	361	6	US-08-233-642A-54	Sequence 54, Appl
24	100	21.9	100	45	US-10-146-492B-77	Sequence 77, Appl
25	100	21.9	100	45	US-10-146-492B-78	Sequence 78, Appl
26	100	21.9	100	45	US-10-146-492B-79	Sequence 79, Appl
27	81.8	17.9	522	17	US-09-252-691-170	Sequence 170, App
28	81.8	17.9	522	17	US-09-252-691C-170	Sequence 170, App
29	81.8	17.9	522	51	US-10-417-886-170	Sequence 170, App
30	43.2	9.5	1083	1	PCT-US01-08631-12549	Sequence 12549, A
C 31	42.6	9.3	37551	81	US-60-245-201-6	Sequence 6, Appli
C 32	42.6	9.3	37551	82	US-60-258-252-2	Sequence 2, Appli
C 33	42.6	9.3	135259	1	PCT-US02-25766-8350	Sequence 8350, Ap
C 34	42.6	9.3	135259	47	US-10-240-425-1585	Sequence 1585, Ap
C 35	42.6	9.3	135259	99	US-60-423-586-24	Sequence 24, Appl
C 36	42.6	9.3	135259	99	US-60-427-194-24	Sequence 24, Appl
37	41.6	9.1	561	17	US-09-252-691-171	Sequence 171, App
38	41.6	9.1	561	17	US-09-252-691C-171	Sequence 171, App
39	41.6	9.1	561	51	US-10-417-886-171	Sequence 171, App
40	39	8.6	1390	43	US-10-017-161-1781	Sequence 1781, Ap
41	39	8.6	1390	48	US-10-292-798-1437	Sequence 1437, Ap
C 42	39	8.6	2730	1	PCT-US01-08631-23586	Sequence 23586, A
43	38.2	8.4	625	43	US-10-029-386-24675	Sequence 24675, A
44	38.2	8.4	32768	77	US-60-213-177-372	Sequence 372, App
45	38.2	8.4	87681	82	US-60-258-275-239	Sequence 239, App

ALIGNMENTS

RESULT 1
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. Coli

US-09-543-407-3

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.9e-130;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
```

RESULT 2
US-60-352-946-1
; Sequence 1, Application US/60352946
; GENERAL INFORMATION:
; APPLICANT: Hultgren, Scott
; APPLICANT: Matt, Chapman
; TITLE OF INVENTION: BACTERIAL MODEL SYSTEM FOR AMYLOID FORMATION
; FILE REFERENCE: WSHU 2059
; CURRENT APPLICATION NUMBER: US/60/352,946
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Escherichia coli
US-60-352-946-1

Query Match 99.3%; Score 453; DB 92; Length 453;
Best Local Similarity 100.0%; Pred. No. 7.5e-129;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
```

```
Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTAC 453
              |||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTAC 453
```

```
RESULT 3
US-60-444-371-1
; Sequence 1, Application US/60444371
; GENERAL INFORMATION:
;  APPLICANT: Hultgren, Scott J
;  APPLICANT:  Matt, Chapman R
;  TITLE OF INVENTION: BACTERIAL MODEL SYSTEM FOR AMYLOID FORMATION
;  FILE REFERENCE: WSHU 2059.1
;  CURRENT APPLICATION NUMBER: US/60/444,371
;  CURRENT FILING DATE:  2003-01-31
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 453
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-60-444-371-1
```

```
  Query Match             99.3%;  Score 453;  DB 101;  Length 453;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-129;
  Matches  453;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTAC 453
              |||||||||||||||||||||||||||||||||
Db        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTAC 453
```

```
RESULT 4
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
;  APPLICANT: NORMARK, Staffan
;  APPLICANT: OLSEN, Arne
;  TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
;  FILE REFERENCE: 012889-081
;  CURRENT APPLICATION NUMBER: US/08/978,878
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: SE 8801723-1
;  EARLIER FILING DATE: 1988-05-06
;  EARLIER APPLICATION NUMBER: US 07/347,189
;  EARLIER FILING DATE: 1989-05-04
;  EARLIER APPLICATION NUMBER: US 07/789,437
;  EARLIER FILING DATE: 1991-11-06
;  EARLIER APPLICATION NUMBER: US 07/970,846
;  EARLIER FILING DATE: 1992-11-03
;  EARLIER APPLICATION NUMBER: US 08/187,865
;  EARLIER FILING DATE: 1994-01-28
;  EARLIER APPLICATION NUMBER: US 08/318,519
;  EARLIER FILING DATE: 1994-10-05
;  EARLIER APPLICATION NUMBER: US 08/495,959
;  EARLIER FILING DATE: 1995-06-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-08-978-878-3
```

```
  Query Match             99.3%;  Score 452.8;  DB 13;  Length 675;
  Best Local Similarity   99.6%;  Pred. No. 9.9e-129;
  Matches  454;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||||||||||||||||||||||||||||||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

```
RESULT 5
US-09-741-873B-3
; Sequence 3, Application US/09741873B
; GENERAL INFORMATION:
;  APPLICANT: Normark, Staffan
;  APPLICANT:  Olsen, Arne
;  TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873B
;  CURRENT FILING DATE:  2003-04-04
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;  LENGTH: 675
;  TYPE: DNA
;  ORGANISM: Escherichia coli
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (83)..(538)
US-09-741-873B-3

 Query Match
 Best Local Similarity   99.3%;  Score 452.8;  DB 31;  Length 675;
 Matches  454;  Conservative   99.6%;  Pred. No. 9.9e-129;
                0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCCGGTAGCGCTCTGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

```
RESULT 6
US-08-978-878-1
; Sequence 1, Application US/08978878
; GENERAL INFORMATION:
;  APPLICANT: NORMARK, Staffan
;  APPLICANT: OLSEN, Arne
;  TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
;  FILE REFERENCE: 012889-081
;  CURRENT APPLICATION NUMBER: US/08/978,878
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: SE 8801723-1
;  EARLIER FILING DATE: 1988-05-06
;  EARLIER APPLICATION NUMBER: US 07/347,189
;  EARLIER FILING DATE: 1989-05-04
;  EARLIER APPLICATION NUMBER: US 07/789,437
;  EARLIER FILING DATE: 1991-11-06
;  EARLIER APPLICATION NUMBER: US 07/970,846
;  EARLIER FILING DATE: 1992-11-03
;  EARLIER APPLICATION NUMBER: US 08/187,865
;  EARLIER FILING DATE: 1994-01-28
;  EARLIER APPLICATION NUMBER: US 08/318,519
;  EARLIER FILING DATE: 1994-10-05
;  EARLIER APPLICATION NUMBER: US 08/495,959
;  EARLIER FILING DATE: 1995-06-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;  LENGTH: 396
;  TYPE: DNA
;  ORGANISM: Escherichia coli
US-08-978-878-1

 Query Match
 Best Local Similarity   86.8%;  Score 396;  DB 13;  Length 396;
 Matches  396;  Conservative   100.0%;  Pred. No. 3e-111;
                0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy         121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy         181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy         241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy         301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy         361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy         421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db         361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

```
RESULT 7
US-09-741-873B-1
; Sequence 1, Application US/09741873B
; GENERAL INFORMATION
;  APPLICANT  Normark, Staffan
;  APPLICANT  Olsen, Arne
;  TITLE OF INVENTION  Fibronectin Binding Protein As Well As Its Preparation
```

```
;  FILE REFERENCE: 012889-084
;  CURRENT APPLICATION NUMBER: US/09/741,873B
;  CURRENT FILING DATE:  2003-04-04
;  PRIOR APPLICATION NUMBER: SE 8801723-1
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 360
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-741-873B-1

  Query Match             78.9%;  Score 360;  DB 31;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 4e-100;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy         121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy         181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy         241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy         301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy         361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360


RESULT 8
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-172

  Query Match             60.7%;  Score 276.8;  DB 17;  Length 477;
  Best Local Similarity   75.4%;  Pred. No. 2.4e-74;
  Matches  344;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy           1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
               |||||||||||| ||||| | ||  ||||||||||||||| | || || || ||| | |||
Db          22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy          61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
               ||||  || || || |  |||||||||          ||  | || ||    |||| ||||
Db          82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy         121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
               ||  | ||    |||| ||||||||||||||| |||||||||| || |||||| ||||
Db         142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy         181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
               || || |||||  | ||||| ||||| |||||||||||||| || ||||| || ||||||
Db         202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy         241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
               ||||| |||||||| |||||||| || ||||||||||    |||   ||||| |||||||||
Db         262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy         301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
               || ||  | || || |||||  ||||| | ||||  || |  || ||||||||||| ||
Db         322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy         361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
               |||||||| || || || |||||||| || ||   |||  | || | |||   ||||||
Db         382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy         421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
               |||||||| |||||||||||||| || |||||||
Db         442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


RESULT 9
US-09-252-691C-172
; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-172

  Query Match             60.7%;  Score 276.8;  DB 17;  Length 477;
  Best Local Similarity   75.4%;  Pred. No. 2.4e-74;
  Matches  344;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy           1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
               |||||||||||| ||||| | ||  ||||||||||||||| | || || || ||| | |||
Db          22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy          61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
               ||||  || || || |  |||||||||          ||  | || ||    |||| ||||
Db          82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy         121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
               ||  | ||    |||| |||||||||||||| |||||||||| || |||||| ||||
Db         142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy         181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
```

```
                 || || |||||  | ||||| ||||| ||||||||||||||| || ||||| || ||||||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
                 ||||| |||||||| |||||||| || |||||||||    |||   ||||| ||||||||||
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
                 || ||  | || || |||||  ||||| | ||||  || |  || ||||||||||| ||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                 |||||||| || || || |||||||| || ||   |||  | || |  |||   ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
                 |||||||| |||||||||||||| || ||||||| |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


RESULT 10
US-10-417-886-172
; Sequence 172, Application US/10417886
; GENERAL INFORMATION:
;  APPLICANT: Keith G. Weinstock et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.135
;  CURRENT APPLICATION NUMBER: US/10/417,886
;  CURRENT FILING DATE:  2003-04-17
;  PRIOR APPLICATION NUMBER: US/09/252,691C
;  PRIOR FILING DATE: 1999-02-18
;  PRIOR APPLICATION NUMBER: US 60/094,145
;  PRIOR FILING DATE: 1998-07-24
;  PRIOR APPLICATION NUMBER: US 60/074,787
;  PRIOR FILING DATE: 1998-02-18
;  NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 172
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-10-417-886-172

  Query Match            60.7%;  Score 276.8;  DB 51;  Length 477;
  Best Local Similarity  75.4%;  Pred. No. 2.4e-74;
  Matches  344;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||| ||||| | ||  |||||||||||||| | || || || ||| | |||
Db         22 ATGAAACTTTTCAAAGTGGCAGTTATTGCAGCAATCGTAGTTTCTGGCAGTGCTTTCGCA 81

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||  || || || |  |||||||||          || | || ||    |||| ||||
Db         82 GGTGCGGTACCACAATTTGGCGGCGGCCACGGTGGTGGCTGGGGTGGCGGCAATAACGGC 141

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||    |||| |||||||||||||||| |||||||||| || ||||| ||||
Db        142 CCTGACTCAACCCTGAGCATTTACCAGTACGGCGGCGGTAACTCCGCGCTTGCTTTGCAG 201

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              || || |||||  | ||||| ||||| |||||||||||||| || ||||| || ||||||
Db        202 ACGGACGCCCGAGATTCTGAATTGACCATTACCCAGCATGGTGGTGGTAACGGCGCAGAT 261

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||| |||||||| |||||||| || |||||||||    |||   ||||| ||||||||||
Db        262 GTTGGCCAGGGCTCTGATGACAGTTCTATCGATCTGCTGCAAAAAGGCTTTGGTAACAGC 321

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || || |||||  ||||| | ||||  || |  || ||||||||||| ||
Db        322 GCCACCATCGACCAATGGAATAGCAAAGACTCTGTTATCAACGTGAAACAGTTCGGCGGC 381

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              |||||||| || || || |||||||| || ||   |||  | || |  |||   ||||||
Db        382 GGCAACGGCGCGGCGGTAGACCAGACAGCGTCCGGCTCAACGGTGACTGTGCACCAGGTT 441

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              |||||||| |||||||||||||| || ||||||| |
Db        442 GGCTTTGGCAACAACGCGACCGCACACCAGTACTGA 477


RESULT 11
US-08-233-642A-56
; Sequence 56, Application US/08233642A
;  GENERAL INFORMATION:
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, S. Karen
;    APPLICANT:  Clouthier, Sharon C.
;    APPLICANT:  Doran, James L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SALMONELLA-
;    TITLE OF INVENTION:  BASED VACCINES
;    NUMBER OF SEQUENCES:  58
;    NUMBER OF SEQUENCES:  -
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,642A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-642A-56

  Query Match            55.8%;  Score 254.4;  DB 6;  Length 456;
  Best Local Similarity  72.4%;  Pred. No. 2e-67;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||    | ||| |||||| ||||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||  ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240
```

```
Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  ||| |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 12
US-09-543-407-1
; Sequence 1, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Salmonella enteritidis
US-09-543-407-1

  Query Match           55.8%;  Score 254.4;  DB 23;  Length 456;
  Best Local Similarity  72.4%;  Pred. No. 2e-67;
  Matches  330;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||   | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  |.||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  ||| |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 13
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match           52.3%;  Score 238.4;  DB 23;  Length 456;
  Best Local Similarity  70.2%;  Pred. No. 1.7e-62;
  Matches  320;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | |||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || ||||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||   | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||    ||||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTATGATCAGCTGGTTACCCGTGTTGTTACC 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                |  ||    |  | ||||| |||||| | |||  ||| |  ||| |||||||
Db        361 CATGAAATGGCACATGCAAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 14
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
```

```
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             51.9%;  Score 236.8;  DB 23;  Length 456;
  Best Local Similarity   70.0%;  Pred. No. 5.4e-62;
  Matches  319;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || |||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||   | ||| |||||| |||||||||  || |||| |||| ||||||||||||
Db        121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGCTCTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| | |||   |    ||||||    ||     | | |    ||| ||
Db        181 AGCGATGCCCGTAAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              |  || |||||  | ||| | ||  | || || ||||| ||   ||| ||| | || |
Db        241 GCAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


RESULT 15
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             50.5%;  Score 230.4;  DB 23;  Length 456;
  Best Local Similarity   69.1%;  Pred. No. 5.2e-60;
  Matches  315;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||| | |||| | ||||||||||||| | || || || ||||||||
Db          1 ATGAAACTTTTAAAAGTGGCAGCATTCGCAGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              || || ||||| || |  |||||||||||||| ||    || |||| |||| |  ||||
Db         61 GGCGTCGTTCCACAATGGGGCGGCGGCGGTAATCATAACGGCGGCGGCAATAGTTCCGGC 120

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||  | ||| | |  ||  ||      | |  |  | |    |||| ||| ||||||
Db        121 CCGGACTATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCACTGCAA 180

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |  |||||||||| |||||   ||| |||||||||   ||     |||| || || |||
Db        181 AGCGATGCCCGTAAATCTGAAACGACCATTACCCAGAGCGGTTATGGTAACGGCGCCGAT 240

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              || ||  |||||  | ||| | ||  | || || |||||  ||    ||| ||| | || |
Db        241 GTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAATAAT 300

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              || ||  | || ||||||||||  ||||| || || || || ||    || | ||| |||
Db        301 GCCACCATCGACCAGTGGAACGCTAAAAACTCCGATATTACTGTCGGCCAATACGGCGGT 360

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
                ||||  ||    ||| | ||||| |||||| | |||  ||| |  |||  |||||||
Db        361 AATAACGCCGCGCTGGTTAATCAGACCGCATCTGATTCCAGCGTAATGGTGCGTCAGGTT 420

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              || ||||| |||||||| || ||| | ||||| |||
Db        421 GGTTTTGGCAACAACGCCACGGCTAACCAGTATTAA 456


Search completed: March 19, 2004, 03:01:41
Job time : 2652.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	452.8	99.3	675	5	US-09-741-873C-3	Sequence 3, Appli
	2	396	86.8	396	5	US-09-741-873C-1	Sequence 1, Appli
c	3	34.2	7.5	79977	6	US-10-765-790-58	Sequence 58, Appl
	4	34.2	7.5	186449	6	US-10-021-698A-715	Sequence 715, App
c	5	33	7.2	439	6	US-10-021-698A-1006	Sequence 1006, Ap
	6	32	7.0	1355	6	US-10-045-674A-593	Sequence 593, App
c	7	30.8	6.8	290892	6	US-10-767-471-10809	Sequence 10809, A
c	8	30.4	6.7	945	6	US-10-767-701-7789	Sequence 7789, Ap
c	9	30.2	6.6	63693	7	US-60-548-091-5635	Sequence 5635, Ap
	10	30.2	6.6	290892	6	US-10-767-471-10809	Sequence 10809, A
c	11	30	6.6	372	6	US-10-021-698A-1217	Sequence 1217, Ap
	12	30	6.6	992	1	PCT-US04-05654-1019	Sequence 1019, Ap
c	13	30	6.6	9278	6	US-10-021-698A-796	Sequence 796, App
	14	29.6	6.5	1203	1	PCT-US04-02242-43	Sequence 43, Appl
	15	29.2	6.4	306	6	US-10-100-683-13342	Sequence 13342, A
	16	29.2	6.4	306	6	US-10-100-683-13343	Sequence 13343, A
	17	29.2	6.4	851	6	US-10-767-795-1598	Sequence 1598, Ap
	18	29.2	6.4	2714	6	US-10-152-372-79	Sequence 79, Appl
	19	29.2	6.4	2776	6	US-10-100-683-3409	Sequence 3409, Ap
c	20	29.2	6.4	4660	6	US-10-788-792-80	Sequence 80, Appl
	21	29	6.4	507	1	PCT-US04-05654-1128	Sequence 1128, Ap
c	22	29	6.4	551	6	US-10-767-701-681	Sequence 681, App
	23	29	6.4	753	6	US-10-767-701-9504	Sequence 9504, Ap
c	24	29	6.4	1350	1	PCT-US04-05654-863	Sequence 863, App
	25	28.8	6.3	6705	5	US-09-032-438C-5	Sequence 5, Appli
	26	28.8	6.3	6705	6	US-10-336-215A-5	Sequence 5, Appli
	27	28.8	6.3	6705	6	US-10-336-219A-5	Sequence 5, Appli
	28	28.8	6.3	6705	6	US-10-340-097B-5	Sequence 5, Appli
	29	28.8	6.3	6819	5	US-09-032-438C-2	Sequence 2, Appli
	30	28.8	6.3	6819	6	US-10-336-215A-2	Sequence 2, Appli
	31	28.8	6.3	6819	6	US-10-336-219A-2	Sequence 2, Appli
	32	28.8	6.3	6819	6	US-10-340-097B-2	Sequence 2, Appli
	33	28.8	6.3	7318	6	US-10-783-528-12	Sequence 12, Appl
	34	28.8	6.3	7488	5	US-09-032-438C-1	Sequence 1, Appli
	35	28.8	6.3	7783	6	US-10-336-215A-1	Sequence 1, Appli
	36	28.8	6.3	7783	6	US-10-336-219A-1	Sequence 1, Appli
	37	28.8	6.3	7783	6	US-10-340-097B-1	Sequence 1, Appli
c	38	28.6	6.3	435	6	US-10-021-698A-1216	Sequence 1216, Ap
	39	28.6	6.3	704	1	PCT-US04-05654-1022	Sequence 1022, Ap
	40	28.6	6.3	860	1	PCT-US04-05654-1018	Sequence 1018, Ap
	41	28.6	6.3	906	1	PCT-US04-05654-1020	Sequence 1020, Ap
	42	28.6	6.3	74279	6	US-10-765-790-16	Sequence 16, Appl
c	43	28.4	6.2	524	6	US-10-767-701-22290	Sequence 22290, A
	44	28.2	6.2	654	6	US-10-767-701-4153	Sequence 4153, Ap
c	45	28.2	6.2	1252	1	PCT-US04-05654-877	Sequence 877, App

ALIGNMENTS

RESULT 1
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match 99.3%; Score 452.8; DB 5; Length 675;
Best Local Similarity 99.6%; Pred. No. 2.4e-145;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||||
Db         83 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTGTGGCA 142

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 202

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        203 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 262

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 322

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 382

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 442

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 502

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        503 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 538
```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 86.8%; Score 396; DB 5; Length 396;
Best Local Similarity 100.0%; Pred. No. 5.4e-126;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 60

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 120

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 180

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 240

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 300

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 360

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        361 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 396
```

RESULT 3
US-10-765-790-58/c
; Sequence 58, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
; LENGTH: 79977
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-58

Query Match 7.5%; Score 34.2; DB 6; Length 79977;
Best Local Similarity 48.2%; Pred. No. 0.61;
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```
Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              |||| || | | | | | | || | | | | || | ||| | | | || | ||
Db      15147 GGTGGTGATGGTGATGATGATGATGGTGATGATGAAGGAGATGGTGATGGTGATGGTGGT 15088

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              || |||       |   |   || | |||| ||| |   || |   | |   |
Db      15087 GATGATGATGATGGTGATGATGGTGAGGATGGTGATGGTGATGATGGAGGAGATGGTGAC 15028

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||   || |   |||     | |   |   ||  || ||| | | ||  |||
Db      15027 GGCGATGGTGGTGATGGTGATCATGATGGTGATGATGGTGATGGTGGTGAGGATGGTGAT 14968

Qy        241 GTTGGTCAGGGCTCAGATG 259
              | |||| | ||   |||||
Db      14967 GGTGGTGATGGTGAAGATG 14949
```

RESULT 4
US-10-021-698A-715
; Sequence 715, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL

; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4044US1
; CURRENT APPLICATION NUMBER: US/10/021,698A
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/211,749
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 6160
; SOFTWARE: Patentin 2.1
; SEQ ID NO 715
; LENGTH: 186449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (40991)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (41047)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (41057)..(41156)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (70192)..(70291)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90076)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90104)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90234)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90262)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90290)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90306)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90313)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90316)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (90344)..(90443)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (93719)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (111786)..(111885)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (129353)..(129452)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (144859)..(144958)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (155513)..(155612)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (166141)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (166388)..(166487)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (172617)..(172716)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (177984)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (178024)..(178123)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (180005)..(180104)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (183290)..(183389)
; OTHER INFORMATION: a, t, c or g
US-10-021-698A-715

Query Match 7.5%; Score 34.2; DB 6; Length 186449;
Best Local Similarity 43.9%; Pred. No. 0.91;
Matches 147; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

```
Qy        46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
             |||   | |   |  | || || |  | |  | || |  |  ||| |    | |||||
Db    148605 GGTGATGATGGTGGTGATGGTGATGATGATGATGGTGATGATGGTGATGCTGATGGTGAT 148664

Qy       106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
             | | || || | |       ||  ||       | | | ||    | ||||  | |      |
Db    148665 GATGATGATGGTGATGGTGATGATGGTGATGATAGTGACTGTGATGGTGATGATGGTGAT 148724

Qy       166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
             |    || |          || ||   || ||| |||       |  |    |  ||  ||
Db    148725 GGTGGTGATGATGGTGATGGTGATGGTGACTGTGATGGTGATGGTGGTGATGGTGATGGT 148784

Qy       226 GGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGT 285
             ||| ||| ||  |||| |||| | |     | |||           |||      |  ||
Db    148785 GGTGATGATGGTGATGATGGTGATGCTGATGGTGATGATGATGGTGATGGTGATGATGGT 148844

Qy       286 GGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTT 345
             |     ||| || | | |  | ||||          ||      ||  |||       | |
Db    148845 GATGATGGTGACTGTGATGGTGATGATGGTGATGGTGGTGGTGATGATGATGGTGATGAT 148904

Qy       346 AAACAGTTCGGTGGTGGCAACGGTGCTGCAGTTGA 380
             |    | ||||||  | |||| ||  | |||
Db    148905 GGTGATGGTGATGGTGGTGATGGTGATGATGGTGA 148939
```

RESULT 5
US-10-021-698A-1006/c

```
; Sequence 1006, Application US/10021698A
; GENERAL INFORMATION:
;  APPLICANT: KEITH, TIM
;  APPLICANT:  LITTLE, RANDALL
;  APPLICANT:  VAN EERDEWEGH, PAUL
;  APPLICANT:  DUPUIS, JOSEE
;  APPLICANT:  DEL MASTRO, RICHARD
;  APPLICANT:  SIMON, JASON
;  APPLICANT:  ALLEN, KRISTINA
;  APPLICANT:  PANDIT, SUNIL
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: 2976-4044US1
;  CURRENT APPLICATION NUMBER: US/10/021,698A
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 6160
;  SOFTWARE: Patentin 2.1
; SEQ ID NO 1006
;    LENGTH: 439
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-021-698A-1006

  Query Match
  Best Local Similarity   47.8%;  Score 33;  DB 6;  Length 439;
  Matches   96;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;
Qy          46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
Db         314 GGTGGTGATGGTGGTGGTGGTGATGGTGGTGATGGTGATGGTGGTGATGGCGATGATGGT 255
Qy         106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
Db         254 GATGATGGTGATGGCGATGATGGTGATAGTGATGGTGATGATGGTGATGGTGATAGTGAT 195
Qy         166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
Db         194 GATAATGGTGATGATGATGGTGATGGTGATAGTGATGATAATGGTGATGATGATGGTGAT 135
Qy         226 GGTAATGGTGCAGATGTTGGT 246
Db         134 GGTAGTGGTGATGGTAGTGGT 114

RESULT 6
US-10-045-674A-593
, Sequence 593, Application US/10045674A
, GENERAL INFORMATION:
,  APPLICANT: LADNER, ROBERT C.
,  APPLICANT:  COHEN, EDWARD H.
,  APPLICANT:  NASTRI, HORACIO G.
,  APPLICANT:  ROOKEY, KRISTIN L.
,  APPLICANT:  HOET, RENE
,  APPLICANT:  HOOGENBOOM, HENDRICUS R. J. M.
,  TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
,  TITLE OF INVENTION:  DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
,  TITLE OF INVENTION:  OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
,  FILE REFERENCE: DYAX/002 CIP2
,  CURRENT APPLICATION NUMBER: US/10/045,674A
,  CURRENT FILING DATE:  2001-10-25
,  PRIOR APPLICATION NUMBER: 06/198,069
,  PRIOR FILING DATE: 2000-04-17
,  PRIOR APPLICATION NUMBER: 09/837,306
,  PRIOR FILING DATE: 2001-04-17
,  NUMBER OF SEQ ID NOS: 635
,  SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 593
,    LENGTH: 1355
,    TYPE: DNA
,    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Description of Artificial Sequence: M13-III
;    OTHER INFORMATION: nucleotide sequence
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1)..(1305)
US-10-045-674A-593

  Query Match
  Best Local Similarity   25.4%;  Score 32;  DB 6;  Length 1355;
  Matches   59;  Conservative   64;  Mismatches  109;  Indels    0;  Gaps    0;
Qy          87 CGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCA 146
Db         663 YGGYAARTTYMGWGAYTGYGCYTTYCAYWSYGGYTTYAAYGARGAYCCWTTYGTYTGYGA 722
Qy         147 GTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGAC 206
Db         723 RTAYCARGGYCARWSKWSYGAYYTRCCKCARCCWCCKGTYAAYGCKGGYGGYGGYWSYGG 782
Qy         207 TATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTC 266
Db         783 YGGWGGYWSYGGYGGYGGYWSYGARGGYGGWGGYWSYGARGGWGGYGGYWSYGGRGGYGG 842
Qy         267 AATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGATCAGTGG 318
Db         843 YWSYGGYWSYGGYGAYTTYGAYTAYGARAARATGGCWAAYGCYAAYAARGGS 894

RESULT 7
US-10-767-471-10809/c
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
;    LENGTH: 290892
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(290892)
;    OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

  Query Match
  Best Local Similarity   46.3%;  Score 30.8;  DB 6;  Length 290892;
  Matches  101;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;
Qy          42 CTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGG 101
Db      234185 CTACATTAGTGATGTTGGTAATGGTGGTGGTGATGATGATGGTGATGCCAATGATTATGA 234126
Qy         102 TGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAA 161
Db      234125 TGATGATGATGGTAATGGTGATGATGGTGATGGTGATAGTTACAGTAGTGATGTTGGTAA 234066
Qy         162 CTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGG 221
Db      234065 TGATAGTGGTGGTGACGACGGTGATGCTAATGATTATGATGATGATGGTGGTGGTAATGG 234006
Qy         222 CGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGATG 259
Db      234005 TGATGATGGTAATGTTGGTGATGATCGTGGCTGTGATG 233968

RESULT 8
```

US-10-767-701-7789/c
; Sequence 7789, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 7789
; LENGTH: 945
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS14531_1
US-10-767-701-7789

Query Match 6.7%; Score 30.4; DB 6; Length 945;
Best Local Similarity 53.3%; Pred. No. 1.5;
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 133 CTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGT 192
Db 569 CTGGAAGTAATCTTGCAAGGCGAGGGTCTCAATGGACTTGTTCTTCAAGCTCTTGATGGC 510

Qy 193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGC 252
Db 509 ATCTATGGAGGCACGTGCACCATCCACGGTTGTGATAATTGGAACCTTGTAGGCGAGGGC 450

RESULT 9
US-60-548-091-5635/c
; Sequence 5635, Application US/60548091
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STROKE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001506
; CURRENT APPLICATION NUMBER: US/60/548,091
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 24433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 63693
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-548-091-5635

Query Match 6.6%; Score 30.2; DB 7; Length 63693;
Best Local Similarity 46.8%; Pred. No. 13;
Matches 95; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
Db 53590 GATGGTTATGGTGATGATGGTGGTGATGGTGGTGGTAGTGGTGGTGGTGATGATGGTGGT 53531

Qy 121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
Db 53530 GGTGATAATGGTAGTGATGATAATGGTGATGGTGGTGGAATGGTGGTGATGATGGTTGAT 53471

Qy 181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
Db 53470 AATGATGGTGATGGTGGTGATGATGGTGATGGTGATGTTGATGGCAGTAATGATGGGGGT 53411

Qy 241 GTTGGTCAGGGCTCAGATGACAG 263
Db 53410 GGTGGTGGTGGTGGTGATGATAG 53388

RESULT 10
US-10-767-471-10809
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
; LENGTH: 290892
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(290892)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

Query Match 6.6%; Score 30.2; DB 6; Length 290892;
Best Local Similarity 49.7%; Pred. No. 26;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 97 GGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGC 156
Db 257220 GGTGGTGGTAATAGTGATGGTGGTGATGGTGGTAATGATGGGGGTGATGGTGTTGATGGT 257279

Qy 157 GGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAG 216
Db 257280 GGTGACGGTGCTGTTGGTGGGGGGAATGATGGTGGTGATAGTAACGGAAATCATACTGAT 257339

Qy 217 CATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGG 251
Db 257340 GCAGGTGGTGGTGATTGTAGTCATGTTGATGATGG 257374

RESULT 11
US-10-021-698A-1217/c
; Sequence 1217, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4044US1
; CURRENT APPLICATION NUMBER: US/10/021,698A
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/211,749
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 6160
; SOFTWARE: Patentin 2.1
; SEQ ID NO 1217
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-021-698A-1217

Query Match 6.6%; Score 30; DB 6; Length 372;
Best Local Similarity 43.1%; Pred. No. 1.3;
Matches 144; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

Qy 62 GTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCC 121
Db 340 GTGATGGTGATGATGATGGTGGTGATAGTGATGGTGATGATGGTGGTGATAGTGATGGTG 281

```
Qy        122 CAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAA 181
              ||  |||      |   |       ||  |||  ||| |   ||    || |     |
Db        280 ATGATGGTGATAGTGATGATGGTGGTTATAGTGATGGTGATGATGGTGGTGATAGTGATG 221

Qy        182 CTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATG 241
              |||||   || |  |     |   |    ||| |   || ||||||  ||||
Db        220 GTGATGATAGTGATGATAGTGATGGTGATGATGATGGTGGTGATAGTGATGGTGATGATG 161

Qy        242 TTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCG 301
              ||||       | |||||  |  |    |||           | |    ||| |    |
Db        160 ATGGTGGTAATACTGATGATGGTGCTGGTGATGATGGTGGTGATAGTGATGGTGATGATG 101

Qy        302 CTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTG 361
              |  | |||| |  |   ||  |   ||  |||       | | |       |||| ||
Db        100 GTGATAGTGATGATGGTGGCGATAGTGATGGTGATGATGGTGGTGATAGTGGTGGTGATG 41

Qy        362 GCAACGGTGCTGCAGTTGACCAGACTGCATCTAA 395
              | |||| |    |||  |   ||    |||
Db         40 ATGATGGTGGTAATAGTGATGATGGTGATAGTAA 7
```

RESULT 12
PCT-US04-05654-1019
; Sequence 1019, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Gutterson, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E
; APPLICANT: Kumimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1019
; LENGTH: 992
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (978)..(984)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (987)..(987)
; OTHER INFORMATION: n is a, c, g, or t
PCT-US04-05654-1019

Query Match 6.6%; Score 30; DB 1; Length 992;
Best Local Similarity 61.5%; Pred. No. 2.1;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy         42 CTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGG 101
              ||  ||  | |  |  | ||||| ||    |   || ||||| |||||   | |||||||
Db        324 CTATGGCGGGGGACGAGGAGGTGGTGGAGGTGGATATGGCGGTGGCGGCGGCTACGGTGG 383

Qy        102 TGGCGGTAATAATAGCGG 119
              |||||||    || | ||
Db        384 TGGCGGTGGCTATGGTGG 401
```

RESULT 13
US-10-021-698A-796/c
; Sequence 796, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEWEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4044US1
; CURRENT APPLICATION NUMBER: US/10/021,698A
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/211,749
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 6160
; SOFTWARE: Patentin 2.1
; SEQ ID NO 796
; LENGTH: 9278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-021-698A-796

Query Match 6.6%; Score 30; DB 6; Length 9278;
Best Local Similarity 46.3%; Pred. No. 6.1;
Matches 99; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```
Qy         46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
              ||| | | |   |  | || || | | |    || |  || | |    | | |||||
Db        359 GGTGGTGATGATGGTGATGATGATGGTGACGGTGGTGATGGTGATGGTGATGATGGTGAT 300

Qy        106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
              | |  | |  | ||       | ||||      |   | |      | ||||  ||| ||  |
Db        299 GATGGTGACGGTGGTGATGGTGCTGATGATGATGGTGATGCTGATGGTGATGGTGACGGT 240

Qy        166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
              |   || |   |  |||||   || |  ||       |  | |  |   ||| ||
Db        239 GATGATGATGGTGACGGTGATGATGGTGATGATGGTGATGGTGATGCTGATGGTGGTGGT 180

Qy        226 GGTAATGGTGCAGATGTTGGTCAGGGCTCAGATG 259
              ||| ||| ||  | || || | | |    ||||
Db        179 GGTGATGATGGTGGTGATGATGATGATGATGATG 146
```

RESULT 14
PCT-US04-02242-43
; Sequence 43, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; CURRENT FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 43
;   LENGTH: 1203
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-43

  Query Match             6.5%;  Score 29.6;  DB 1;  Length 1203;
  Best Local Similarity   56.0%;  Pred. No. 3.2;
  Matches   56;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy         36 CGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCA 95
              |||   | |||   ||   |      | ||||  ||| || | | || ||||   | | |
Db        228 CGTGAACACCGTGCGCCTGCGCCTGTGGGTTGACCCTAAGGATGCCGACGGCAACACCTA 287

Qy         96 CGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAGCTG 135
              ||| || ||| ||||||||  ||      |    || ||||
Db        288 CGGCGGCGGCAGTAATAATCTCGAAAACACGATTGCGCTG 327


RESULT 15
US-10-100-683-13342
; Sequence 13342, Application US/10100683
; GENERAL INFORMATION:
;  APPLICANT: Rosen, et al.
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS900
;  CURRENT APPLICATION NUMBER: US/10/100,683
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: US 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: US 60/043,576
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,845
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,599
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,664
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: US 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: US 60/056,892
;  PRIOR FILING DATE: 1997-08-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 13468
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13342
;   LENGTH: 306
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-100-683-13342

  Query Match             6.4%;  Score 29.2;  DB 6;  Length 306;
  Best Local Similarity   52.5%;  Pred. No. 2.3;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        128 CTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATG 187
              || || |||      |  ||  ||  |||      | | |||| ||| |    || ||  ||
Db        124 CTCAGGTGAGTTCCTCTCACCACCCTGGGCAGCAGTGTGCAGTTGGTTACAAAGCTCTTG 183

Qy        188 CCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTC 247
               |  |   ||||  ||||   |   | | ||| |||    |  || || || |||   ||
Db        184 TCTTTTGTTCTGGATTGAAAGTCCGCAAACATAGCGTTTATGCTGTTGAAGCTGTGAATC 243

Qy        248 AG 249
              ||
Db        244 AG 245


Search completed: March 19, 2004, 03:10:00
Job time : 57.1445 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaaacttttaaaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
c	1	453.4	99.4	680	12	BJ618688	BJ618688 BJ618688
c	2	60.8	13.3	558	14	CD374421	CD374421 TNWbmfC2H
c	3	43.4	9.5	607	12	BJ348812	BJ348812 BJ348812
	4	38	8.3	282	29	AL758610	AL758610 Arabidops
c	5	37.2	8.2	540	12	BM382763	BM382763 MEST554-D
c	6	37.2	8.2	771	28	BZ997124	BZ997124 PUGGV50TB
	7	37.2	8.2	797	28	BZ997127	BZ997127 PUGGV50TD
	8	37.2	8.2	818	29	CG200536	CG200536 PUIJS94TD
c	9	37	8.1	789	28	BZ115869	BZ115869 CH230-255
	10	36	7.9	1171	13	BX342087	BX342087 BX342087
c	11	35.8	7.9	675	14	CB642466	CB642466 OSJNEb02K
c	12	35.8	7.9	774	29	CNS02APC	AL188841 Tetraodon
	13	35.8	7.9	935	10	BE572585	BE572585 601329541
	14	35.6	7.8	435	10	BE604165	BE604165 WHE1413-1
	15	35.6	7.8	498	10	BE604467	BE604467 WHE1413-1
c	16	35.6	7.8	523	12	BM380389	BM380389 MEST519-C
	17	35.6	7.8	537	28	B64870	B64870 CIT-HSP-201
	18	35.6	7.8	642	28	BZ421645	BZ421645 hz32f09.b
c	19	35.6	7.8	652	12	BM379271	BM379271 MEST503-B
c	20	35.6	7.8	684	29	CC616249	CC616249 OGWEH75TV
	21	35.6	7.8	769	29	CG101505	CG101505 PUIEO88TD
c	22	35.6	7.8	793	29	CG364848	CG364848 OG3AZ85TV
c	23	35.6	7.8	839	28	BZ969045	BZ969045 PUGFX86TD
c	24	35.6	7.8	854	29	CG142634	CG142634 PUFQE81TD
	25	35.6	7.8	867	29	CG142632	CG142632 PUFQE81TB
	26	35.6	7.8	880	28	BZ571017	BZ571017 msh2_168.
	27	35.6	7.8	921	29	CG364838	CG364838 OG3AZ85TH
c	28	35.6	7.8	926	29	CC627745	CC627745 OGVBG10TV
	29	35.6	7.8	926	29	CG211571	CG211571 OGXDF50TV
	30	35.6	7.8	950	28	CC327866	CC327866 OGUEO87TH
	31	35.6	7.8	970	29	CC691748	CC691748 OGOBQ41TV
c	32	35.6	7.8	1245	28	BZ578681	BZ578681 msh2_5964
c	33	35.4	7.8	244	29	CE219585	CE219585 tigr-gss-
	34	35.2	7.7	379	12	BJ266281	BJ266281 BJ266281
	35	35.2	7.7	634	12	BJ266096	BJ266096 BJ266096
	36	35.2	7.7	708	12	BJ272995	BJ272995 BJ272995
	37	35.2	7.7	1251	14	CD509711	CD509711 CDA97-G08
c	38	35	7.7	629	12	BJ404131	BJ404131 BJ404131
	39	34.8	7.6	685	12	BJ249080	BJ249080 BJ249080
c	40	34.6	7.6	582	12	BM276856	BM276856 952012H03
	41	34.6	7.6	586	10	AW759708	AW759708 sl47a01.y
c	42	34.6	7.6	701	14	CD436924	CD436924 EL01N0365
	43	34.6	7.6	818	14	CD098848	CD098848 AGENCOURT
	44	34.6	7.6	978	29	CNS00K6G	AL077284 Drosophil
c	45	34.2	7.5	306	14	R55792	R55792 yg89h05.s1

ALIGNMENTS

RESULT 1
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibb.ac.jp.

```
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match             99.4%;  Score 453.4;  DB 12;  Length 680;
  Best Local Similarity   99.6%;  Pred. No. 1.5e-114;
  Matches  454;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAAACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        593 ATGAAACTTTTAAAAGTAGCAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCA 534

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 474

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        473 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTNTGCACTTGCTCTGCAA 414

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 354

Qy        241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGC 294

Qy        301 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 GCTACTCTTGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGT 234

Qy        361 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 GGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAACGTGACTCAGGTT 174

Qy        421 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        173 GGCTTTGGTAACAACGCGACCGCTCATCAGTACTAA 138


RESULT 2
CD374421/c
LOCUS       CD374421                 558 bp    mRNA    linear   EST 19-AUG-2003
DEFINITION  TNWbmfC2H3T3 Wuchereria bancrofti microfilaria cDNA (SAW95SjL-WbMf)
            Wuchereria bancrofti cDNA clone TNWbmfC2H3 5', mRNA sequence.
ACCESSION   CD374421
VERSION     CD374421.1  GI:31229939
KEYWORDS    EST.
SOURCE      Wuchereria bancrofti
  ORGANISM  Wuchereria bancrofti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Wuchereria.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
  TITLE     Toward the characterization of the Wuchereria bancrofti
            microfilarial transcriptome with comparisons to those of Brugia
            malayi and Onchocerca volvulus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Thomas B. Nutman
            Laboratory of Parasitic Diseases
            NIH
            Building 4 Room 126, Bethesda, MD 20892-0425, USA
            Email: tnutman@niaid.nih.gov

            Seq primer: pBluescript T3.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Wuchereria bancrofti"
                     /mol_type="mRNA"
                     /db_xref="taxon:6293"
                     /clone="TNWbmfC2H3"
                     /dev_stage="microfilaria"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="Wuchereria bancrofti microfilaria cDNA
                     (SAW95SjL-WbMf)"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 85,000 microfilariae
                     isolated from the blood of an infected individual from
                     Guyana, South America and converted to double-stranded
                     cDNA using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNA pol I. The library has 2.9 x 10E5
                     independent recombinants and the average insert size is
                     ~1kb. The library was constructed by Sandra J. Laney. The
                     library is available from Dr.S.A.Williams, email:
                     genome@smith.edu."
ORIGIN

  Query Match             13.3%;  Score 60.8;  DB 14;  Length 558;
  Best Local Similarity   77.5%;  Pred. No. 6.8e-06;
  Matches   93;  Conservative    0;  Mismatches   25;  Indels    2;  Gaps    2;

Qy         63 TGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAAT-AGCGGCC 121
              |||||    |     |    |||| |||||||||| ||||||||||||||||||| |||||||
Db        174 TGTTGGCACCTGCTGGCGCGGTGGCGGTAACCNCGGTGGTGGCGGTAATAATAAGCGGCC 115

Qy        122 CAAATTCTGAGCTGAACATTTACC-AGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              ||||||||  |||||    ||  || ||  ||||||||| | |||||| ||||||||| ||
Db        114 CAAATTCTANGCTGACATTTANCCAAGANCGGTGGCGGNANCTCTGCNCTTGCTCTGNAA 55


RESULT 3
BJ348812/c
LOCUS       BJ348812                 607 bp    mRNA    linear   EST 06-MAR-2002
DEFINITION  BJ348812 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda34o11 3', mRNA sequence.
ACCESSION   BJ348812
VERSION     BJ348812.1  GI:19219319
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda34o11"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match              9.5%;  Score 43.4;  DB 12;  Length 607;
```

```
  Best Local Similarity   47.7%;  Pred. No. 0.47;
  Matches   95;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy         61 GGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGC 120
              | || || || | |  | | ||| |  | | |  ||  || ||| | | || ||   | |
Db        351 GATGGTGATCATNACCATGACGGTGATGATNATGACCATGATGGTGATGATGATNATGAC 292

Qy        121 CCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAA 180
              |    | |||     |   | ||||  |||||  | | |   |   | || |    |
Db        291 CATGGTGATGATGATNATNATGACCATGACGGTGATGATGATGATNACCATGATGNTGAT 232

Qy        181 ACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGAT 240
              |||||  || | ||  |||      |       |   ||  |    | ||  ||  |||
Db        231 GNTGATGACCATGACGNTGATGATNATNACCATGANGNTGATGATNATNATCATGGTGAT 172

Qy        241 GTTGGTCAGGGCTCAGATG 259
              | || | ||   |  ||||
Db        171 GATGATNATNACCANGATG 153


RESULT 4
AL758610
LOCUS       AL758610                282 bp    DNA     linear   GSS 18-JUN-2002
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-160E03-013216,
            genomic survey sequence.
ACCESSION   AL758610
VERSION     AL758610.1  GI:21496958
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 282)
  AUTHORS   Rosso,M., Strizhov,N., Li,Y. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion close to or within gene At3g28590. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .282
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-160E03-013216"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
ORIGIN

  Query Match             8.3%;  Score 38;  DB 29;  Length 282;
  Best Local Similarity   46.8%;  Pred. No. 10;
  Matches  109;  Conservative    0;  Mismatches  123;  Indels    1;  Gaps    1;

Qy         13 AAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGTTGTTCCT 72
              |||  || |   | |||||   || |  | | |  ||     |
Db          1 AAATAAGGAAATAATGCAGATTTCCTTATTTTCCTTATTTTACANNNNNNNNNNNNNNNN 60

Qy         73 CAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTGAG 132
                  || |  | ||||   || |||| ||||  ||||||  |   |||      |||
Db         61 NNNNNNNATGGTGATGATAACTGTGGCGGTGCCGGTGGTAATAGTGAAACAACAGTGGAG 120

Qy        133 CTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGT 192
              ||   ||      | | |||| || ||||  |||   |||          || || | |
Db        121 GTGGCGATGGTGGCG-ATGGTGACGATAACAGTGCCGGTGCCGGTGGTGATGGTGACGGC 179

Qy        193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGG 245
              |||  || |   ||  |  | |   || || |  |||||||  |  | |||
Db        180 AACAATGGCGGTACCGGTGACGATGGTGACGACAATAATGGTGGTGCCGGTGG 232


RESULT 5
BM382763/c
LOCUS       BM382763                  540 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  MEST554-D07.univ ISUM6 Zea mays cDNA clone MEST554-D07 3', mRNA
            sequence.
ACCESSION   BM382763
VERSION     BM382763.1  GI:18181553
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Wen,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.
  TITLE     Expressed Sequence Tags from B73 Maize: various stages and tissues
            including seedlings treated with a variety of hormones
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Patrick S. Schnable
            Schnable Laboratory
            Iowa State University
            G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
            Tel: 515-294-0975
            Fax: 515-294-2299
            Email: schnable@iastate.edu
            Individual basecall and confidence value were assigned using the
            Phred software(http://www.phrap.org/). Overall sequence quality
            assessment and vector trimming was conducted using the Lucy
            software (version 1.16s, http://www.tigr.org/softlab/). Lucy
            parameters were set to ensure an overall trimmed quality of 97.5%
            or better without any vector fragments in the chosen high quality
            region of each sequence. Low-quality bases between the poly-T and
            the high-quality region were replaced with N's to serve as spacers
            using a Perl program (est_process.pl), written by Dr. Hui-Hsien
            Chou.
            PCR PRimers
            FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
            BACKWARD: primer Sp6 (ATT TAG GTG ACA CTA TAG)
            Seq primer: universal (GTA AAA CGA CGG CCA GT)
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="MEST554-D07"
                     /tissue_type="mixed"
                     /lab_host="DH10B"
```

/clone_lib="ISUM6"
/note="Vector: pSlip7 (4.43 kb); Site_1: EcoRI; Site_2: NotI; Tissue samples were collected and partially pooled prior to RNA extraction. First-strand cDNAs were prepared from 21 individual pools of oligo-dT selected mRNAs by priming with 21 different NotI oligo-dT tag primers (5'-AACTGGAAGAATTCGCGGCCGCNNNNNNTTTTTTTTTTTTTTTT-3'). Distinguishable 'bar code' tags, (N)6, were used for each separate first-strand cDNA synthesis. Hence, these bar code tags can be used to identify the mRNA pool from which a particular cDNA clone was derived. The 'bar code' tags associated with specific tissue sources are: ATACGC--Germinated seeds and seedlings (1, 2, 8, 11 DAG); ACTGGC--Mixed mature tissues (17, 21, 38, 69, 77 DAG); CACAGC--Kernels (3, 5, 10, 15, 20, 25, 30, DAP); TAACCC--Adventious roots (65 DAG); CAGGCG--Tassels (3-39 cm, 53 and 56 DAG); AGGTAC--Immature ears (0.2-3.0 cm, 53, 56, 59 DAG); TGAGCG--Husks (73 DAG); GACCAC--Silks; AATCGG--unpollinated first ears; CTAAGG--ear shanks; GAAGAG--etiolated seedlings; AGTGAG--callus; GTGGAC--Cycloheximide-treated callus; GTCACC--Anaerobic treated seedlings; CGTCCA--NAA (a-Naphthalene acetic acid)-treated seedlings; GATGCC--Kinetin-treated seedlings; AAGACC--ACPC (1-aminocyclopropane-1-carboxylix acid)-treated seedlings; GCCTCA--Brassinolide-treated seedlings; CTAGCC--ABA (Abscisic acid)-treated seedlings; TACGGA--GA (Gibberellic acid)-treated seedlings; GCAGGA--JA (Jasmonic acid)-treated seedlings; Equal amounts of first-strand cDNA from each reaction were combined and used as template for DNA PolI-catalyzed 2nd strand synthesis. After the addition of EcoRI adaptors, ds-cDNAs were digested with NotI. Molecules between 0.5 and 2.0 kb were directionally cloned into the EcoRI and NotI sites of the pSlip7 expression vector. Plasmid DNA isolated from the library was digested with NotI to remove empty vector clones. Linear DNAs from 5.4 to 7 kb were gel purified and ligated at low concentration to promote recircularization. Ligation products were precipitated and transformed into DH10B host cells. The complexity of the resulting library was 1.2 x 10X6."

ORIGIN

Query Match 8.2%; Score 37.2; DB 12; Length 540;
Best Local Similarity 55.4%; Pred. No. 23;
Matches 72; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy         45 CGGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGG 104
              |||    | ||| |  ||  ||| |  ||    ||||||||||||      |||||| ||
Db        410 CGGGTCAGGTCTCGGCGGCCTTGGTGGTCTCGGCGGCGGCGGCGGCCTTGGCGGTGGCGG 351

Qy        105 CGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTC 164
              |||   | | |||||  ||   || |  |||  |   ||     |  |||||||  |
Db        350 CGGCCTTGGTGGCGGCTCAGGATCGGGTCTGGGCGGCAACGGTGGCCTTGGCGGTGGCAG 291

Qy        165 TGCACTTGCT 174
              ||  || | |
Db        290 TGGGCTCGGT 281
```

RESULT 6
BZ997124/c
LOCUS BZ997124 771 bp DNA linear GSS 25-MAR-2003
DEFINITION PUGGV50TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa381I04, genomic survey sequence.
ACCESSION BZ997124
VERSION BZ997124.1 GI:29240541
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 771)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUGGV50TD
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .771
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa381I04"
/clone_lib="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high CoT selected genomic DNA library"
ORIGIN

Query Match 8.2%; Score 37.2; DB 28; Length 771;
Best Local Similarity 50.6%; Pred. No. 27;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy        193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGC 252
              |||  ||   |      || ||| | | || || || ||||  | ||    ||   |
Db        374 AACACTTCGTCAGTAATAACCGAACGTGTCGTCGATAATAATTCAATGCTTCACCGTTAA 315

Qy        253 TCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGAT 312
              | |   ||   || | |  | ||| | |||  | |||||||  | | | || ||| ||
Db        314 TAATCCAACGTTCCACTAGTAATAACCAAGCGTTTCATCGGTAATAACCCAACGCTTCAT 255

Qy        313 CAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTG 370
              | ||   |   ||   |||   | | |      |||  ||||  |||   ||| | |
Db        254 CGGTAATAATCGAACGTTTCGTGAGTAATAACCCAACGCTTCGCCGGTAATAACCGAG 197
```

RESULT 7
BZ997127
LOCUS BZ997127 797 bp DNA linear GSS 25-MAR-2003
DEFINITION PUGGV50TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa381I04, genomic survey sequence.
ACCESSION BZ997127
VERSION BZ997127.1 GI:29240544
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 797)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUGGV50TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .797

```
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa381I04"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             8.2%;  Score 37.2;  DB 28;  Length 797;
  Best Local Similarity   50.6%;  Pred. No. 28;
  Matches   90;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGC 252
              ||| ||   |     || ||| | | || || || ||||  | ||    ||   |
Db        505 AACACTTCGTCAGTAATAACCGAACGTGTCGTCGATAATAATTCAATGCTTCACCGTTAA 564

Qy        253 TCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGAT 312
              | |    ||    || | | | ||| | ||| | ||||||| | | | || ||| ||
Db        565 TAATCCAACGTTCCACTAGTAATAACCAAGCGTTTCATCGGTAATAACCCAACGCTTCAT 624

Qy        313 CAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTG 370
              | ||   |    ||  |||   | | |       |||  ||||  |||   ||| | |
Db        625 CGGTAATAATCGAACGTTTCGTGAGTAATAACCCAACGCTTCGCCGGTAATAACCGAG 682


RESULT 8
CG200536
LOCUS       CG200536                818 bp    DNA     linear   GSS 21-AUG-2003
DEFINITION  PUIJS94TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0599O19,
            genomic survey sequence.
ACCESSION   CG200536
VERSION     CG200536.1  GI:34091597
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa0599O19"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             8.2%;  Score 37.2;  DB 29;  Length 818;
  Best Local Similarity   50.6%;  Pred. No. 28;
  Matches   90;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        193 AACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGC 252
              ||| ||   |     || ||| | | || || || ||||  | ||    ||   |
Db        277 AACACTTCGTCAGTAATAACCGAACGTGTCGTCGATAATAATTCAATGCTTCACCGTTAA 336

Qy        253 TCAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGAT 312
              | |    ||    || | | | ||| | ||| | ||||||| | | | || ||| ||
Db        337 TAATCCAACGTTCCACTAGTAATAACCAAGCGTTTCATCGGTAATAACCCAACGCTTCAT 396

Qy        313 CAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTGGTGGCAACGGTG 370
              | ||   |    ||  |||   | | |       |||  ||||  |||   ||| | |
Db        397 CGGTAATAATCGAACGTTTCGTGAGTAATAACCCAACGCTTCGCCGGTAATAACCGAG 454


RESULT 9
BZ115869/c
LOCUS       BZ115869                789 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-255D7.TVB CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-255D7, genomic survey sequence.
ACCESSION   BZ115869
VERSION     BZ115869.1  GI:23756816
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-255D7.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 255  row: D  column: 7
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-255D7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             8.1%;  Score 37;  DB 28;  Length 789;
  Best Local Similarity   45.8%;  Pred. No. 31;
  Matches  127;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy         46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
              ||| | | |   |  | || || | |    | || || || ||| |  | | || ||
Db        680 GGTGGTGGTGGTGATGCTGATGATGTTGCTGATGGTGGTGGTGGTGATGATGCTGATGAT 621

Qy        106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
              | |  | || | |       ||  ||       |     |  |    |||| | |     |
Db        620 GTTGCTGATGGTGCTGGTGATGATGCTGATGATGAGGATGATGGTGGTGATGATGGTGAT 561
```

```
Qy        166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
              |  ||||        ||||||   ||     |||      ||  | |  |   ||| ||
Db        560 GGTGATGCTGCTGCTGCTGATGGTGGTGGTGGTGATGCTGCTGCTGCTGATGGTGGTGGT 501

Qy        226 GGTAATGGTGCAGATGTTGGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAACGT 285
              ||| ||| ||| | || || |   ||    | |||      |    ||| |  |  | ||
Db        500 GGTGATGCTGCTGCTGCTGATGGTGGTGGTGGTGATGATGCTGATGATGTTGCTGATGGT 441

Qy        286 GGCTTCGGTAACAGCGCTACTCTTGATCAGTGGAACG 322
              ||    ||| |    |||| |  |||| |   | | |
Db        440 GGTGGTGGTGATGATGCTAATGCTGATGATGAGGATG 404


RESULT 10
BX342087
LOCUS       BX342087               1171 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  BX342087 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DK009YH08 5-PRIME, mRNA sequence.
ACCESSION   BX342087
VERSION     BX342087.1  GI:30308886
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1171)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2235.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DK009DD04QP1
            &cluster=2235.f. Contact : Feng Liang Email : fliang@lifetech.com
            URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DK009DD04QP1.
FEATURES             Location/Qualifiers
     source          1. .1171
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DK009YH08"
                     /cell_type="HELA CELLS COT 25-NORMALIZED"
                     /cell_line="HELA"
                     /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             7.9%;  Score 36;  DB 13;  Length 1171;
  Best Local Similarity   12.1%;  Pred. No. 72;
  Matches   22;  Conservative   95;  Mismatches   65;  Indels    0;  Gaps    0;

Qy         71 CTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGTAATAATAGCGGCCCAAATTCTG 130
              | | : : ||:::: |:| :  ||:  : : :|::  : :  : ::: ||:::::: :::
Db        947 CCCCKSMGGGSKSMAGMGMKCKCCMGCKCKMKGKSCMKGMRCMGMBMMGCMMMMMKCMKV 1006

Qy        131 AGCTGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCC 190
               |  :  ::: : :: : : :::: :::::::  : :::   : :::: :::|::|:::
Db       1007 KGGCKGGMMKSBGMMTKMMKKKKKMMKKKMMMMAKCMMMMMMMMMWKMTMMMTKMTKMMW 1066

Qy        191 GTAACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGG 250
              :: :  :: ::  :   ::: ::| |::  :::::| : :::   | |  | :|  | ::
Db       1067 KKKMKGYHHDMMMKKKSMKKKMMCKGBMCMKBKBBGCDGMKKMMCCCGCCGCKGCCCCSS 1126

Qy        251 GC 252
              ::
Db       1127 SS 1128


RESULT 11
CB642466/c
LOCUS       CB642466               675 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb02K15.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb02K15 3', mRNA sequence.
ACCESSION   CB642466
VERSION     CB642466.1  GI:29637457
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 02  row: K  column: 15
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1. .675
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb02K15"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 14;  Length 675;
  Best Local Similarity   59.2%;  Pred. No. 63;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          6 ACTTTTAAAAGTAGAAGCAATTGCAGCAATCGTATTCTCCGGTAGCGCTCTGGCAGGTGT 65
              ||||  | | |  |||||    || ||   ||   | || |||    |||  || |||||
Db        626 ACTTCCACACGGTGAAGCTGCCGCTGCTGCCGCCGTCGGGTGAACCTCCTGCGGGTGT 567

Qy         66 TGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGGT 108
                ||||  |    |   | ||||   ||||||| |||||||||
Db        566 GCTTCCGGACGGTGTACGTGGCGAGCACCACGGCGGTGGCGGT 524


RESULT 12
CNS02APC/c
LOCUS       CNS02APC               774 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            252I15 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL188841
VERSION     AL188841.1  GI:7826945
KEYWORDS    GSS; genome survey sequence.
```

```
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 774)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="252I15"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG252AE08SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 29;  Length 774;
  Best Local Similarity   48.7%;  Pred. No. 67;
  Matches   97;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          46 GGTAGCGCTCTGGCAGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGC 105
               ||| | | |   | | || || | | | | || | |  ||| |  | | ||||||
Db         445 GGTGGTGATGGTGGTGATGATGATGATGATGATGGTGATGATGGTGATGATGATGGTGGT 386

Qy         106 GGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACTCT 165
               ||| || || | ||     ||  |||        | |  |  | ||||| |||     |
Db         385 GGTGATGATGGTGGTGATGATGGTGATGATGGTGGTGATGATGATGGTGGTGGTGGTGAT 326

Qy         166 GCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCGGC 225
               |    || |         |||||   ||      |||       | |     |  |||| ||
Db         325 GATGATGGTGGTGGTGGTGATGATGGTGGTGATGATGGTGGTGGTGGTGATGATGGTGGT 266

Qy         226 GGTAATGGTGCAGATGTTG 244
               | | ||||||  |||| ||
Db         265 GATGATGGTGGTGATGATG 247


RESULT 13
BE572585
LOCUS       BE572585                935 bp    mRNA    linear   EST 15-AUG-2000
DEFINITION  601329541F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3603347 5',
            mRNA sequence.
ACCESSION   BE572585
VERSION     BE572585.1  GI:9816305
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8792  row: a  column: 12
            High quality sequence stop: 679.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3603347"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
ORIGIN

  Query Match             7.9%;  Score 35.8;  DB 10;  Length 935;
  Best Local Similarity   53.1%;  Pred. No. 73;
  Matches   76;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         104 GCGGTAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCGGTAACT 163
               || | ||  | | | | |||||  |     ||  | ||  | | | |    |  ||| ||
Db         485 GCTGAAAAGAGATCTGGCCAAAGACATCACTTCAGATACATCTGGAGACTTTCGGAAAGC 544

Qy         164 CTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTATTACCCAGCATGGCG 223
               ||   ||||||||    | | || ||||      |||||| | | |  ||  || | |
Db         545 CTTTGCTTGCTCTTGCCAAGGGTGACCGTTGTCAGGACTTGAGTGTGAATCAAGATTGGG 604

Qy         224 GCGGTAATGGTGCAGATGTTGGT 246
               | | ||  | |||     || ||
Db         605 GTGATACAGATGCCAGGCTTTGT 627


RESULT 14
BE604165
LOCUS       BE604165                435 bp    mRNA    linear   EST 21-AUG-2000
DEFINITION  WHE1413-1416_I13_I13ZS Wheat drought stressed leaf cDNA library
            Triticum aestivum cDNA clone WHE1413-1416_I13_I13, mRNA sequence.
ACCESSION   BE604165
VERSION     BE604165.1  GI:9861435
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Han,P.S., Hsia,C.C.,
            Kang,Y., Lazo,G.R., Miller,R., Nguyen H.T., Rausch,C.J.,
            Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
```

genomes - Drought stressed leaf cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

FEATURES Location/Qualifiers
source 1. .435
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="TAM W101"
/db_xref="taxon:4565"
/clone="WHE1413-1416_I13_I13"
/tissue_type="Leaf"
/dev_stage="Full tillering stage"
/lab_host="E. coli SOLR"
/clone_lib="Wheat drought stressed leaf cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were given a gradual stress down to 65% and 78% RWC at Texas Tech University (D. Zhang in HT Nguyen lab). Total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close), at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Query Match 7.8%; Score 35.6; DB 10; Length 435;
Best Local Similarity 50.0%; Pred. No. 57;
Matches 89; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy        134 TGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTA 193
              | ||  |  ||   || ||    ||| |   | || |||   ||||| | | ||||   |
Db         10 TCAAATTGGACGTCTATGGAAATGGTGAAGATTCAGTTGAAGTGCAATCAGCTGCCAAGA 69

Qy        194 ACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCT 253
              |    || |||| | ||| |   |||   ||| |  ||  |||||||| |  ||
Db         70 AGCTGGATTTGAATCTTAACTTCCATAAAGGCCGAGATCATGCAGATGATTCTCTTCATG 129

Qy        254 CAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGA 311
                |  |   |  | | || |   |||  ||    || ||| |||  |||
Db        130 GGTACAAGGTTTTTGTAAACCCGAGCATTAGTGATGTCCTCTGCACCGCCACTGCTGA 187
```

RESULT 15
BE604467
LOCUS BE604467 498 bp mRNA linear EST 21-AUG-2000
DEFINITION WHE1413-1416_B07_B07ZS Wheat drought stressed leaf cDNA library Triticum aestivum cDNA clone WHE1413-1416_B07_B07, mRNA sequence.
ACCESSION BE604467
VERSION BE604467.1 GI:9861825
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 498)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Nguyen H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
TITLE The structure and function of the expressed portion of the wheat genomes - Drought stressed leaf cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

FEATURES Location/Qualifiers
source 1. .498
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="TAM W101"
/db_xref="taxon:4565"
/clone="WHE1413-1416_B07_B07"
/tissue_type="Leaf"
/dev_stage="Full tillering stage"
/lab_host="E. coli SOLR"
/clone_lib="Wheat drought stressed leaf cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were given a gradual stress down to 65% and 78% RWC at Texas Tech University (D. Zhang in HT Nguyen lab). Total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close), at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Query Match 7.8%; Score 35.6; DB 10; Length 498;
Best Local Similarity 50.0%; Pred. No. 61;
Matches 89; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy        134 TGAACATTTACCAGTACGGTGGCGGTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTA 193
              | ||  |  ||   || ||    ||| |   | || |||   ||||| | | ||||   |
Db         10 TCAAATTGGACGTCTATGGAAATGGTGAAGATTCAGTTGAAGTGCAATCAGCTGCCAAGA 69

Qy        194 ACTCTGACTTGACTATTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCT 253
              |    || |||| | ||| |   |||   ||| |  ||  |||||||| |  ||
Db         70 AGCTGGATTTGAATCTTAACTTCCATAAAGGCCGAGATCATGCAGATGATTCTCTTCATG 129

Qy        254 CAGATGACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCTTGA 311
                |  |   |  | | || |   |||  ||    || ||| |||  |||
Db        130 GGTACAAGGTTTTTGTAAACCCGAGCATTAGTGATGTCCTCTGCACCGCCACTGCTGA 187
```

Search completed: March 18, 2004, 20:11:15
Job time : 2037.55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 2309.32 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		456	100.0	4680	1	ECCSGABDG	X90754 E.coli csgG
2		456	100.0	10346	1	AE000205	AE000205 Escherich
3		456	100.0	15047	1	D90741	D90741 Escherichia
4		454.4	99.6	1711	1	AF275733	AF275733 Escherich
5		454.4	99.6	10190	1	AE005315	AE005315 Escherich
6		454.4	99.6	327773	1	AP002554	AP002554 Escherich
7		451.2	98.9	10370	1	AE015131	AE015131 Shigella
8		451.2	98.9	306358	1	AE016759	AE016759 Escherich
9		449.6	98.6	292504	1	AE016981	AE016981 Shigella
10		356.8	78.2	2889	1	CSP515700	AJ515700 Citrobact
11		332.8	73.0	2067	1	SEU43280	U43280 Salmonella
12		332.8	73.0	5103	1	STAJ2301	AJ002301 Salmonell
13		332.8	73.0	22411	1	AE008749	AE008749 Salmonell
14		331.2	72.6	254050	1	AL627269	AL627269 Salmonell
15	c	331.2	72.6	301983	1	AE016840	AE016840 Salmonell
16		323	70.8	2920	1	CFR515701	AJ515701 Citrobact
17		275	60.3	1212	1	ECO131756	AJ131756 Escherich
18		266.4	58.4	1048	1	STAGFBA	AJ000514 Salmonell
19		258.2	56.6	2883	1	ESA515702	AJ515702 Enterobac
20		51.2	11.2	301214	1	AE016786	AE016786 Pseudomon
21		49.4	10.8	246435	2	AC096956	AC096956 Rattus no
22	c	47.6	10.4	253001	3	AE014834	AE014834 Plasmodiu
23	c	43.4	9.5	234090	2	AC111458	AC111458 Rattus no
24		43.2	9.5	182148	2	AC119000	AC119000 Rattus no
25		43	9.4	3732	3	PFASA27	M10129 Plasmodium
26		42.6	9.3	210554	2	AC133982	AC133982 Rattus no
27	c	41.4	9.1	10709	1	AE015532	AE015532 Shewanell
28	c	40.4	8.9	172742	2	AC136466	AC136466 Rattus no
29	c	40.2	8.8	242510	2	AC128507	AC128507 Rattus no
30	c	39.8	8.7	160460	2	AC024374	AC024374 Homo sapi
31	c	39.8	8.7	161536	2	AC024318	AC024318 Homo sapi
32		39.8	8.7	179693	9	AL591856	AL591856 Human DNA
33		39.8	8.7	192016	2	AC026540	AC026540 Homo sapi
34	c	39.6	8.7	7218	6	I66494	I66494 Sequence 14
35	c	39.6	8.7	257757	3	AE014837	AE014837 Plasmodiu
36		39.2	8.6	74524	9	AL512508	AL512508 Human DNA
37	c	39.2	8.6	190950	2	AL359639	AL359639 Homo sapi
38	c	38.6	8.5	50820	1	AE014163	AE014163 Streptoco
39		38.6	8.5	310850	1	AP005142	AP005142 Streptoco
40		38.2	8.4	4860	6	AR320671	AR320671 Sequence
41	c	38.2	8.4	186401	2	AC097009	AC097009 Pan trogl
42	c	38.2	8.4	242730	2	AC099408	AC099408 Pan trogl
43	c	37.8	8.3	2000	6	AX655393	AX655393 Sequence
44	c	37.8	8.3	31521	3	CEY47H10A	AL132848 Caenorhab
45	c	37.8	8.3	141489	10	AL591376	AL591376 Mouse DNA

ALIGNMENTS

RESULT 1
ECCSGABDG
LOCUS ECCSGABDG 4680 bp DNA linear BCT 07-JUL-2002
DEFINITION E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION X90754
VERSION X90754.1 GI:1147558
KEYWORDS csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene; orfC gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Hammar,M., Arnqvist,A., Bian,Z., Olsen,A. and Normark,S.
TITLE Expression of two csg operons is required for production of

```
            fibronectin- and congo red-binding curli polymers in Escherichia
            coli K-12
  JOURNAL   Mol. Microbiol. 18 (4), 661-670 (1995)
  MEDLINE   96414468
   PUBMED   8817489
REFERENCE   2  (bases 1 to 4680)
  AUTHORS   Hammar,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
            Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="23.15 minutes"
     source          1. .3140
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="MC4100"
                     /db_xref="taxon:562"
     source          3141. .4680
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="W3110"
                     /db_xref="taxon:562"
     repeat_region   <1. .>4680
                     /insertion_seq="IS2 (partial)"
     gene            complement(133. .966)
                     /gene="csgG"
     CDS             complement(133. .966)
                     /gene="csgG"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62277.1"
                     /db_xref="GI:1147559"
                     /db_xref="GOA:P52103"
                     /db_xref="SWISS-PROT:P52103"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene.           complement(993. .1409)
                     /gene="csgF"
     CDS             complement(993. .1409)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62278.1"
                     /db_xref="GI:1147560"
                     /db_xref="SWISS-PROT:P52104"
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(1434. .1823)
                     /gene="csgE"
     CDS             complement(1434. .1823)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="assembly /transport component in curli
                     production"
                     /protein_id="CAA62279.1"
                     /db_xref="GI:1147561"
                     /db_xref="SWISS-PROT:P52105"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(1828. .2478)
                     /gene="csgD"
     CDS             complement(1828. .2478)
                     /gene="csgD"
                     /note="necessary for transcription of the csgBA operon"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAA62280.1"
                     /db_xref="GI:1147562"
                     /db_xref="GOA:P52106"
                     /db_xref="SWISS-PROT:P52106"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     misc_feature    complement(2626)
                     /note="transcriptional start site for csgDEFG operon"
     misc_feature    3140
                     /note="transcriptional start site for csgBA operon"
     gene            3233. .3688
                     /gene="csgB"
     CDS             3233. .3688
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleation component of curlin monomers"
                     /protein_id="CAA62281.1"
                     /db_xref="GI:1147563"
                     /db_xref="GOA:P39828"
                     /db_xref="SWISS-PROT:P39828"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            3729. .4184
                     /gene="csgA"
     CDS             3729. .4184
                     /gene="csgA"
                     /note="major subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin/CsgA protein"
                     /protein_id="CAA62282.1"
                     /db_xref="GI:1147564"
                     /db_xref="GOA:P28307"
                     /db_xref="SWISS-PROT:P28307"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGNHGGGGNNSGPNSE
                     LNIYQYGGGNSALALQTDARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSAT
                     LDQWNGKNSEMTVKQFGGGNGAAVDQTASNSSVNVTQVGFGNNATAHQY"
     sig_peptide     3729. .3788
                     /gene="csgA"
     gene            4243. .4575
                     /gene="orfC"
     CDS             4243. .4575
                     /gene="orfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="CAA62283.1"
                     /db_xref="GI:1147565"
                     /db_xref="SWISS-PROT:P52107"
                     /translation="MNTLLLLAALSSQITFNTTQQGDVYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
ORIGIN

  Query Match           100.0%;  Score 456;  DB 1;  Length 4680;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-116;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3233 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 3292

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3293 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 3352

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3353 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 3412

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3413 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 3472

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3473 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 3532

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3533 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 3592

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3593 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 3652

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       3653 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 3688
```

RESULT 2
AE000205
LOCUS AE000205 10346 bp DNA linear BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 95 of 400 of the complete genome.
ACCESSION AE000205 U00096
VERSION AE000205.1 GI:1787265
KEYWORDS .
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10346)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 10346)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 10346)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 10346)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES Location/Qualifiers
source 1. .10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
repeat_region complement(<1. .48)
/note="IS3"
promoter <1. .5
/note="factor Sigma70; predicted +1 start at 1094689"
promoter 5. .32
/note="factor Sigma70; predicted +1 start at 1094716"
gene 69. .392
/gene="b1028"
CDS 69. .392
/gene="b1028"
/function="orf; Unknown"
/note="o107; This 107 aa ORF is 27 pct identical (0 gaps) to 58 residues of an approx. 1888 aa protein FAS2_CANAL SW: P43098"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74113.1"
/db_xref="GI:1787266"
/translation="MEDGVQAMRDYLAGLDIASPEHQVLMNVTAKSEVAPSIIKENLS LHLTHTVKWTESLDTFLNMPTPVAFLEISNKPYLGNMLNDFAGVDQQRVMHCRKAFSD AKVFK"
promoter 275. .299
/gene="b1028"
/note="factor Sigma32; predicted +1 start at 1094983"
gene 389. .1375
/gene="ycdU"
/note="synonym: b1029"
CDS 389. .1375
/gene="ycdU"
/function="orf; Unknown"
/note="o328; This 328 aa ORF is 19 pct identical (4 gaps) to 191 residues of an approx. 288 aa protein Y320_MYCGE SW: P47562"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"

```
                     /translation="MIPDYLTFIRFQDKRNLIYIYAIGLILIGFYWKNAGFTFPSEDI
                     GVVSGILALVLYNFIFDLKAYWAYKCVTKNIDFSWFKKKQNHKIELPLTQPLVAGFLS
                     LIMLSAMSWGLYQLLPSLYALFLISLLGPLVIFLLFRMIRTSYVKQVAISVAKKVKYK
                     SLTRYVLLSVCISTVVNLLTISPLRNSDSFVTEGQWLTFKSIIALLILCGVVLAINLF
                     FLRFSKRYAFLGRLFLQEIDLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVL
                     ATLVEWRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
protein_bind         1253. .1280
                     /gene="ycdU"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="DeoR predicted site"
protein_bind         1253. .1280
                     /gene="ycdU"
                     /note="No predicted promoter"
                     /bound_moiety="DeoR predicted site"
promoter             1337. .1365
                     /gene="ycdU"
                     /note="factor Sigma70; predicted +1 start at 1096049"
gene                 1494. .1745
                     /gene="b1030"
CDS                  1494. .1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
                     to81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="MGGGTRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHC
                     VHGLSNWLSPVVDPRSGLLIPPVCAIHEKKARTFVRALL"
gene                 1513. .1926
                     /gene="ycdV"
                     /note="synonym: b1031"
CDS                  1513. .1926
                     /gene="ycdV"
                     /function="putative structure; Not classified"
                     /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
                     to47 residues of an approx. 80 aa protein RS37_YEAST SW:
                     P05759"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosomal protein"
                     /protein_id="AAC74116.1"
                     /db_xref="GI:1787269"
                     /translation="MCNIRKKSPYFRTSSSLNMAVRGGLTRFARPAGSPLTAFTVCPT
                     GCRQLSTPGRGFSSPRCVQYTKKKPVLSYELFFKYGGEGGIDSLRSPCGQPAHCVHGL
                     SNWLSPVVEPRSGLLIPPVCAIYEKKARTFVRALL"
gene                 complement(2111. .2198)
                     /gene="serX"
                     /note="synonym: b1032"
tRNA                 complement(2111. .2198)
                     /gene="serX"
                     /product="tRNA-Ser"
                     /note="anticodon: GGA; CG Site No. 17878"
                     /function="RNA; tRNA"
promoter             2269. .2296
                     /note="factor Sigma70; predicted +1 start at 1096980"
promoter             2338. .2365
                     /note="factor Sigma70; predicted +1 start at 1097049"
promoter             2355. .2383
                     /note="factor Sigma70; predicted +1 start at 1097067"
gene                 2393. .3370
                     /gene="ycdW"
                     /note="synonym: b1033"
CDS                  2393. .3370
                     /gene="ycdW"
                     /function="putative enzyme; Not classified"
                     /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
                     to 177 residues of an approx. 312 aa protein YPRA_CORGL
                     SW: P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74117.1"
                     /db_xref="GI:1787270"
                     /translation="MRINFAQRFSGVSMDIIFYHPTFDTQWWIEALRKAIPQARVRAW
                     KSGDNDSADYALVWHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFR
                     LEDTGMGEQMQEYAVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVL
                     GSKVAQSLQTWRFPLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETV
                     GIINQQLLEKLPDGAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESP
                     LWQHPRVTITPHVAAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
promoter             3278. .3308
                     /gene="ycdW"
                     /note="factor Sigma70; predicted +1 start at 1097992"
promoter             3383. .3410
                     /note="factor Sigma70; predicted +1 start at 1098094"
promoter             3396. .3423
```

```
  Query Match           100.0%;  Score 456;  DB 1;  Length 10346;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-116;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8497 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 8556

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8557 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 8616

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8617 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 8676

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8677 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 8736

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8737 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 8796

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8797 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 8856

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8857 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 8916

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       8917 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 8952
```

```
RESULT 3
D90741
LOCUS       D90741                15047 bp    DNA     linear   BCT 25-DEC-2002
DEFINITION  Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION   D90741 AB001340
VERSION     D90741.1  GI:1651509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
            csgA; ycdE; cls; nov; mdoG.
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
```

```
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
   PUBMED   8905232
REFERENCE   2
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 15047)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                      Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .15047
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:83333"
                     /map="23.7-24.0 min"
                     /clone="Kohara clone #231"
                     /note="Nucleotide position 1097101-1112147 from the
                     initiation site of ThrA (0 min.).-This clone is from
                     Kohara lambda miniset library."
     CDS             <1. .276
                     /note="unnamed protein product; ORF_ID:o230#5"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35822.1"
                     /db_xref="GI:4062605"
                     /translation="DLFFSSENALSTFFAKPLWLRLFILLVIEVMWITLVSVLATLVE
                     WRIWFEAYFLLCYVPCLIYYFFYCRFLWHNDFMMACDMYFRWGHFNK"
     CDS             1152. .2090
                     /note="ORF_ID:o231#1
                     similar to SwissProt Accession Number P45637"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical 33.0 kd protein in proB-proA
                     intergenic region."
                     /protein_id="BAA35823.1"
                     /db_xref="GI:4062606"
                     /translation="MDIIFYHPTFDTQWWIEALRKAIPQARVRAWKSGDNDSADYALV
                     WHPPVEMLAGRDLKAVFALGAGVDSILSKLQAHPEMLNPSVPLFRLEDTGMGEQMQEY
                     AVSQVLHWFRRFDDYRIQQNSSHWQPLPEYHREDFTIGILGAGVLGSKVAQSLQTWRF
                     PLRCWSRTRKSWPGVQSFAGREELSAFLSQCRVLINLLPNTPETVGIINQQLLEKLPD
                     GAYLLNLARGVHVVEDDLLAALDSGKVKGAMLDVFNREPLPPESPLWQHPRVTITPHV
                     AAITRPAEAVEYISRTIAQLEKGERVCGQVDRARGY"
     CDS             2145. .2882
                     /note="unnamed protein product; ORF_ID:o231#2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35824.1"
                     /db_xref="GI:4062607"
                     /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                     EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                     PVFAPHDKATNTQAMIATIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                     SFLHSRKGSEDNCREVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPPER
                     ILNVSPRRLLNFLESRGMAPIAEFADL"
     CDS             2906. .3460
                     /note="ORF_ID:o231#3
                     similar to PIR Accession Number F64035"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein HI1543"
                     /protein_id="BAA35825.1"
                     /db_xref="GI:4062608"
                     /translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ
                     DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEGATEAEVRAFLSERGMPL
                     ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW
                     RTMAPLTRDAISAMWDELEEDSEE"
     CDS             3514. .4053
                     /note="unnamed protein product; ORF_ID:o231#4"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA35826.1"
                     /db_xref="GI:4062609"
                     /translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL
                     SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG
                     IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLIFGYAMKNSGLW
                     LAARSAKTAHREQEIKNKA"
     gene            complement(4117. .4950)
                     /gene="csgG"
     CDS             complement(4117. .4950)
                     /gene="csgG"
                     /note="ORF_ID:o231#5
                     similar to SwissProt Accession Number P52103"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgG precursor."
                     /protein_id="BAA35827.1"
                     /db_xref="GI:4062610"
                     /translation="MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     VGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE
                     S"
     gene            complement(4977. .5393)
                     /gene="csgF"
     CDS             complement(4977. .5393)
                     /gene="csgF"
                     /note="ORF_ID:o231#6
                     similar to SwissProt Accession Number P52104"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgF precursor."
                     /protein_id="BAA35828.1"
                     /db_xref="GI:4062611"
```

```
                     /translation="MRVKHAVVLLMLISPLSWAGTMTFQFRNPNFGGNPNNGAFLLNS
                     AQAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIV
                     DIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF"
     gene            complement(5418. .5807)
                     /gene="csgE"
     CDS             complement(5418. .5807)
                     /gene="csgE"
                     /note="ORF_ID:o231#7
                     similar to SwissProt Accession Number P52105"
                     /codon_start=1
                     /transl_table=11
                     /product="Curli production assembly/transport component
                     CsgE precursor."
                     /protein_id="BAA35829.1"
                     /db_xref="GI:4062612"
                     /translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
                     RAFSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDFEKTVVFAL
                     IQTEEALNRRQINQALLSTGDLAHDEF"
     gene            complement(5812. .6462)
                     /gene="csgD"
     CDS             complement(5812. .6462)
                     /gene="csgD"
                     /note="ORF_ID:o231#8
                     similar to SwissProt Accession Number P52106"
                     /codon_start=1
                     /transl_table=11
                     /product="Probable csgAB operon transcriptional regulatory
                     protein."
                     /protein_id="BAA35830.1"
                     /db_xref="GI:4062613"
                     /translation="MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQ
                     RSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWP
                     HINGVFYSMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            7217. .7672
                     /gene="csgB"
     CDS             7217. .7672

  Query Match           100.0%;  Score 456;  DB 1;  Length 15047;
  Best Local Similarity  100.0%;  Pred. No. 2.6e-116;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7217 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 7276

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7277 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 7336

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7337 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 7396

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7397 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 7456

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7457 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 7516

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7517 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 7576

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7577 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 7636

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       7637 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 7672
```

```
RESULT 4
AF275733
LOCUS       AF275733               1711 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
            CsgA protein (csgA) genes, complete cds.
ACCESSION   AF275733
VERSION     AF275733.1  GI:14039399
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Mutations in the csgD promoter associated with variations in curli
            expression in certain strains of Escherichia coli O157:H7
  JOURNAL   Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
  MEDLINE   21218556
   PUBMED   11319125
REFERENCE   2  (bases 1 to 1711)
  AUTHORS   Uhlich,G.A., Keen,J.E. and Elder,R.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
            Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES             Location/Qualifiers
     source          1. .1711
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="43895 Red Variant"
                     /serotype="O157:H7"
                     /db_xref="taxon:562"
     gene            757. .1212
                     /gene="csgB"
     CDS             757. .1212
                     /gene="csgB"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="AAK53211.1"
                     /db_xref="GI:14039400"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1253. .1711
                     /gene="csgA"
     CDS             1253. .1711
                     /gene="csgA"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="AAK53212.1"
                     /db_xref="GI:14039401"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
ORIGIN

  Query Match           99.6%;  Score 454.4;  DB 1;  Length 1711;
  Best Local Similarity  99.8%;  Pred. No. 7.6e-116;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 816

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 876

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        877 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 936

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        937 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 996

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        997 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 1056

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1057 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 1116

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 1176

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1177 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 1212


RESULT 5
AE005315
LOCUS       AE005315               10190 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
            of 155.
ACCESSION   AE005315 AE005174
VERSION     AE005315.1  GI:12514572
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10190)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10190
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            617. .1072
                     /gene="csgB"
                     /note="synonym: Z1675"
     CDS             617. .1072
                     /gene="csgB"
                     /function="structural component; Cell exterior
                     constituents: Surface structures"
                     /note="Residues 1 to 151 of 151 are 100.00 pct identical
                     to residues 1 to 151 of 151 from Escherichia coli K-12
                     Strain MG1655: B1041"
                     /codon_start=1
                     /transl_table=11
                     /product="minor curlin subunit precursor, similar ro CsgA"
                     /protein_id="AAG55787.1"
                     /db_xref="GI:12514573"
                     /translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAVNELSKSSFN
                     QAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYIDQAGSANDA
                     SISQGAYGNTAMIIQKGSGNKANITQYGTQKTAIVVQRQSQMAIRVTQR"
     gene            1113. .1571
                     /gene="csgA"
                     /note="synonym: Z1676"
     CDS             1113. .1571
                     /gene="csgA"
                     /function="regulator; Cell exterior constituents: Surface
                     structures"
                     /note="Residues 1 to 152 of 152 are 96.71 pct identical to
                     residues 1 to 151 of 151 from Escherichia coli K-12 Strain
                     MG1655: B1042"
                     /codon_start=1
                     /transl_table=11
                     /product="curlin major subunit, coiled surface structures;
                     cryptic"
                     /protein_id="AAG55788.1"
                     /db_xref="GI:12514574"
                     /translation="MKLLKVAAIAAIVFSGSALAGVVPQYGGGGGNHGGGGNNSGPNS
                     ELNIYQYGGGNSALALQADARNSDLTITQHGGGNGADVGQGSDDSSIDLTQRGFGNSA
                     TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTQVGFGNNATAHQY"
     gene            1630. .1962
                     /gene="csgC"
                     /note="synonym: Z1677"
     CDS             1630. .1962
                     /gene="csgC"
                     /function="putative structure; Not classified"
                     /note="Residues 1 to 110 of 110 are 98.18 pct identical to
                     residues 1 to 110 of 110 from Escherichia coli K-12 Strain
                     MG1655: B1043"
                     /codon_start=1
                     /transl_table=11
                     /product="putative curli production protein"
                     /protein_id="AAG55789.1"
                     /db_xref="GI:12514575"
                     /translation="MNALLLLAALSSQITFNTTQQGDMYTIIPEVTLTQSCLCRVQIL
                     SLREGSSGQSQTKQEKTLSLPANQPIALTKLSLNISPDDRVKIVVTVSDGQSLHLSQQ
                     WPPSSEKS"
     gene            2083. .2394
                     /gene="Z1678"
     CDS             2083. .2394
                     /gene="Z1678"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 103 of 103 are 98.05 pct identical to
                     residues 1 to 103 of 103 from Escherichia coli K-12 Strain
                     MG1655: B1044"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55790.1"
                     /db_xref="GI:12514576"
                     /translation="MFRPPFLNSLMLGSLFPPFIAIAGSTAQGGVIHFYGQIVEPACDV
                     STQSSPVEMNCPQNGSVPGKTYSSKALMSGNVKNAQIASVKVQYLDKQKKLAVMNIEY
                     N"
     gene            2489. .3022
                     /gene="Z1679"
     CDS             2489. .3022
                     /gene="Z1679"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 177 of 177 are 100.00 pct identical
                     to residues 1 to 177 of 177 from Escherichia coli K-12
                     Strain MG1655: B1045"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyprotein"
                     /protein_id="AAG55791.1"
```

```
                /db_xref="GI:12514577"
                /translation="MKTRIHVVQGDITKLAVDVIVNAANPSLMGGGGVDGAIHRAAGP
                ALLDACLKVRQQQGDCPTGHAVITLAGDLPAKAVVHTVGPVWRGGEQNEDQLLQDAYL
                NSLRLVAANSYTSVAFPAISTGVYGYPRAAAAEIAVKTVSEFITRHALPEQVYFVCYD
                EENAHLYERLLTQQGDE"
gene            2964. .4445
                /gene="ymdC"
                /note="synonym: Z1680"
CDS             2964. .4445
                /gene="ymdC"
                /function="putative enzyme; Not classified"
                /note="Residues 1 to 493 of 493 are 99.59 pct identical to
                residues 1 to 493 of 493 from Escherichia coli K-12 Strain
                MG1655: B1046"
                /codon_start=1
                /transl_table=11
                /product="putative synthase"
                /protein_id="AAG55792.1"
                /db_xref="GI:12514578"
                /translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPGQCGLFPL
                EKSLDAFAARYRLAEMSEHTLDVQYYIWQDDMSGRLLFSALLAAAKRGVRVRLLLDDN
                NTPGLDDILRLLDSHPRIEVRLFNPFSFRLLRPLGYITDFSRLNRRMHNKSFTVDGVV
                TLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWYCKSVSPLQQVLDV
                PEGEMADRIELPASWHNDAMTHRYLRKMESSPFINHLVDGTLPLIWAKTRLLSDDPAK
                GEGKAKRHSLLPQRLFDIMGSPSERIDIISSYFVPTRAGVAQLLRMVRKXVKIAILTN
                SLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREQSSTLHDRGITGNSGASLHAKT
                FSIDGKTVFIGSFNFDPRSTLLNTEMGFVIESETLAQLIDKRFIQSQYDAAWQLRLDR
                WGRINWVDRHAKKEIVLKKEPATSFWKRVMVRLASILPVEWLL"
gene            complement(4453. .5610)
                /gene="ymdD"
                /note="synonym: Z1681"
CDS             complement(4453. .5610)
                /gene="ymdD"
                /function="orf; Unknown function"
                /note="Residues 1 to 385 of 385 are 99.22 pct identical to
                residues 1 to 385 of 385 from Escherichia coli K-12 Strain
                MG1655: B1047"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG55793.1"
                /db_xref="GI:12514579"
                /translation="MNPVPAQREYFXDSIRAWLMLLGIPFHISLIYSSHTWHVNSAEP
                SLWLTLFNDFIHSFRMQVFFVISGYFSYMLFLRYPLKKWWKVRVERVGIPMLTAIPLL
                TLPQFIMLQYVKGKAESWPGLSLYDKYNTLAWELISHLWFLLVLVVMTTLCVWIFKRI
                RNNLENSDKTNKKFSMVKLSVIFLCLGIGYAVIRRTIFIVYPPILSNGTFNFIVMQTL
                FYLPFFILGALAFIFPHLKALFTTPSRGCTLAAALAFVAYLLNQRYGSGDAWMYETES
                VITMVLGLWMVNVVFSFGHRLLNFQSARVTYFVNASLFIYLVHHPLTLFFGAYITPHI
                TSNWLGFLCGLIFVVGIAIILYEIHLRIPLLKFLFSGKPVVKRENDKAPAR"
gene            6003. .7538
                /gene="mdoG"
                /note="synonym: Z1683"
CDS             6003. .7538
                /gene="mdoG"
                /function="enzyme; Osmotic adaptation"
                /note="Residues 1 to 511 of 511 are 99.80 pct identical to
                residues 1 to 511 of 511 from Escherichia coli K-12 Strain
                MG1655: B1048"
                /codon_start=1
                /transl_table=11
                /product="periplasmic glucans biosynthesis protein"
                /protein_id="AAG55794.1"
                /db_xref="GI:12514580"
                /translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAQSLAGKGYEAPK
                SNLPSVFRDMKYADYQQIQFNHDKAYWNNLKTPFKLEFYHQGMYFDTPVKINEVTATA
                VKRIKYSPDYFTFGDVQHDKDTVKDLGFAGFKVLYPINSKDKNDEIVSMLGASYFRVI
                GAGQVYGLSARGLAIDTALPSGEEFPRFKEFWIERPKPTDKRLTIYALLDSPRATGAY
                KFVVMPGRDTVVDVQSKIYLRDKVGKLGVAPLTSMFLFGPNQPSPANNYRPELHDSNG
                LSIHAGNGEWIWRPLNNPKHLAVSSFSMENPQGFGLLQRGRDFSRFEDLDDRYDLRPS
                AWVTPKGEWGKGSVELVEIPTNDETNDNIVAYWTPDQLPEPGKEMNFKYTITFSRDED
                KLHAPDNAWVQQTRRSTGDVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDTPVTAQTS
                IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDAKKTTEMRAALVNADQTLSETWSYQLP
                ANE"
```

```
    gene            7531. .10074
                    /gene="mdoH"
                    /note="synonym: Z1684"
    CDS             7531. .10074
                    /gene="mdoH"
                    /function="enzyme; Osmotic adaptation"

  Query Match           99.6%;  Score 454.4;  DB 1;  Length 10190;
  Best Local Similarity  99.8%;  Pred. No. 7.4e-116;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 676

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 736

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 796

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 856

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 916

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        917 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 976

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 1036

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db       1037 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 1072
```

```
RESULT 6
AP002554
LOCUS       AP002554               327773 bp    DNA     linear   BCT 21-DEC-2002
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION   AP002554 BA000007
VERSION     AP002554.1  GI:13360491
KEYWORDS    .
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
   PUBMED   10734605
REFERENCE   2
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
```

```
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
   PUBMED   11108008
REFERENCE   3
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
   PUBMED   11111050
REFERENCE   4
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
   PUBMED   11258796
REFERENCE   5  (bases 1 to 327773)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .327773
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            150. .2258
                     /gene="ECs1032"
     CDS             150. .2258
                     /gene="ECs1032"
                     /note="similar to YCBY_ECOLI gi|1787181 percent identity
                     99 in 702 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative oxidoreductase"
                     /protein_id="BAB34455.1"
                     /db_xref="GI:13360492"
                     /translation="MNSLFASTARGLEELLKTELENLGAVECQVVQGGVHFKGDTRLV
                     YQSLMWSRLASRIMLPLGECKVYSDLDLYLGVQAINWTEMFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRDAPDIRVNVWLHKETASIALDLSGDGLH
                     LRGYRDRAGIAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMLATDRAPGL
                     HRGRWGFSGWAQHDEAIWQEVKAEAQTRARKGLABYSSHFYGSDSDARVIQRARTNAR
                     LAGIGELITFEVKDVAQLTNPLPKEPYGTVLSNPPYGERLDSEPALIALHSLLGRIMK
                     NQFGGWNLSLFSASPDLLSCLQLRADKQYKAKNGPLDCVQKNYHVAESTPDSKPAMVA
                     EDYANRLRKNLKKFEKWARQEGIECYRLYDADLPEYNVAVDRYADWVVVQEYAPPKTI
                     DAHKARQRLFDIIAATISVLGIAPNKLVLKTRERQKGKNQYQKLGEKGEFLEVTEYNA
                     HLWVNLTDYLDTGLFLDHRIARRMLGQMSKGKDFLNLFSYTGSATVHAGLGGARSTTT
                     VDMSRTYLEWAERNLRLNGLTGRAHRLIQADCLAWLREANEQFDLIFIDPPTFSNSKR
                     MEDAFDVQRDHLALMKDLKRLLRAGGTIMFSNNKRGFRMDLDGLAKLGLKAQEITQKT
                     LSQDFARNRQIHNCWLITAA"
     gene            2270. .4177
                     /gene="ECs1033"
     CDS             2270. .4177
                     /gene="ECs1033"
                     /note="similar to UUP_ECOLI gi|1787182 percent identity 99
                     in 635 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative ATP-binding component of a transport
                     system"
                     /protein_id="BAB34456.1"
                     /db_xref="GI:13360493"
                     /translation="MSLISMHGAWLSFSDAPLLDNAELHIEDNERVCLVGRNGAGKST
                     LMKILNREQGLDDGRIIYEQDLIVARLQQDPPRNVEGSVYDFVAEGIEEQAEYLKRYH
                     DISRLVMNDPSEKNLNELAKVQEQLDHHNLWQLENRINEVLAQLGLDPNVALSSLSGG
                     WLRKAALGRALVSNPRVLLLDEPTNHLDIETIDWLEGFLKTFNGTIIFISHDRSFIRN
                     MATRIVDLDRGKLVTYPGNYDQYLLEKEEALRVEELQNAEFDRKLAQEEVWIRQGIKA
                     RRTRNEGRVRALKAMRRERGERREVMGTAKMQVEEASRSGKIVFEMEDVCYQVDGKQL
                     VKDFSAQVLRGDKIALIGPNGCGKTTLLKLMLGQLQADSGRIHVGTKLEVAYFDQHRA
                     ELDPDKTVMDNLAEGKQEVMVNGKPRHVLGYLQDFLFHPKRAMTPVRALSGGERNRLL
                     LARLFLKPSNLLILDEPTNDLDVETLELLEELIDSYQGTVLLVSHDRQFVDNTVTECW
                     IFEGGGKIGRYVGGYHDARGQQEQYVALKQPAVKKNEEPAAPKAETVKRSSSKLSYKL
                     QRELEQLPQLLEDLEAKLEALQTQVADASFFSQPHEQTQKVLADMAAAEQELEQAFER
                     WEYLEALKNGG"
     gene            4307. .5560
                     /gene="ECs1034"
     CDS             4307. .5560
                     /gene="ECs1034"
                     /note="similar to PQIA_ECOLI gi|1787183 percent identity
                     99 in 417 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="paraquat-inducible protein A"
                     /protein_id="BAB34457.1"
                     /db_xref="GI:13360494"
                     /translation="MCEHHHAAKHILCSQCDMLVALPRLEHGQKAACPRCGTTLTVAW
                     DAPRQRPTAYALAALFMLLLSNLFPFVNMNVAGVTSEITLLEIPGVLFSEDYASLGTF
                     FLLFVQLVPAFCLITILLLVNRAELPVRLKEQLARVLFQLKTWGMAEIFLAGVLVSFV
                     KLMAYGSIGVGSSFLPWCLFCVLQLRAFQCVDRRWLWDDIAPMPELRQPLKPGVTGIR
                     QGLRSCSCCTAILPADEPVCPRCGTKGYVRRRNSLQWTLALLVTSIMLYLPANILPIM
                     VTDLLGSKMPSTILAGVILLWSEGSYPVAAVIFLASIMVPTLKMIAIAWLCWDAKGHG
                     KRDSERMHLIYEVVEFVGRWSMIDVFVIAVLSALVRMGGLMSIYPAMGALMFALVVIM
                     TMFSAMTFDPRLSWDRQPESEHEES"
     gene            5565. .7205
                     /gene="ECs1035"
     CDS             5565. .7205
                     /gene="ECs1035"
                     /note="similar to PQIB_ECOLI gi|1787184 percent identity
                     99 in 546 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="paraquat-inducible protein B"
                     /protein_id="BAB34458.1"
                     /db_xref="GI:13360495"
                     /translation="MESNNGEAKIQKVKNWSPVWIFPIVTALIGAWVLFYHYSHQGPE
                     VTLITANAEGIEGGKTTIKSRSVDVGVVESATLADDLTHVEIKARLNSGMEKLLHKDT
                     VFWVVKPQIGREGISGLGTLLSGVYIELQPGAKGSKMDKYDLLDSPPLAPPDAKGIRV
                     VLDSKKAGQLSPGDPVLFRGYRVGSVETSTFDTQKRNISYQLFINAPYDRLVTSNVRF
                     WKDSGIAVDLTSAGMRVEMGSLTTLLSGGVSFDVPEGLDLGQPVAPKTAFVLYDDQKS
                     IQDSLYTDHIDYLMFFKDSVRGLQPGAPVEFRGIRLGTVSKVPFFAPNMRQTFNDDYR
                     IPVLIRIEPERLKMQLGENADVVEHLGELLKRGLRGSLKTGNLVTGALYVDLDFYPNT
                     PAITGIREFNGYQIIPTVSGGLAQIQQRLMEALDKINKLPLNPMIEQATSTLSESQRT
                     MKNLQTTLDSMNKILASQSMQQLPTDMQSTLRELNRSMQGFQPGSAAYNKMVADMQRL
                     DQVLRELQPVLKTLNEKSNALVFEAKDKKDPEPKRAKQ"
     gene            7217. .7765
                     /gene="ECs1036"
     CDS             7217. .7765
                     /gene="ECs1036"
                     /note="similar to YMBA_ECOLI gi|1787185 percent identity
                     98 in 182 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB34459.1"
                     /db_xref="GI:13360496"
```

```
                     /translation="MTIAALWLAGCSSGEINKNYYQLPVVQSGTQSTASQGNRLLWVE
                     QVAVPDYLAGNGVVYQTSDVKYVIANNNLWASPLDQQLRNTLVANLSTQLPGWVVASQ
                     PLGSAQDTLNVTVTEFNGRYDGKVIVSGEWLLNHQGQLIKRPFRLEGVQTQDGYDEMV
                     KVLAGVWSQEAASIAQEIKRLP"
     gene            8021. .8188
                     /gene="ECs1037"
     CDS             8021. .8188
                     /gene="ECs1037"
                     /note="identical to RMF_ECOLI gi|1787186 (Conserved in
                     E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ribosome modulation factor"
                     /protein_id="BAB34460.1"
                     /db_xref="GI:13360497"
                     /translation="MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGW
                     REAMADRVVMA"
     gene            complement(8258. .8776)
                     /gene="ECs1038"
     CDS             complement(8258. .8776)
                     /gene="ECs1038"
                     /note="similar to FABA_ECOLI gi|1787187 percent identity
                     100 in 172 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11

  Query Match             99.6%;  Score 454.4;  DB 1;  Length 327773;
  Best Local Similarity   99.8%;  Pred. No. 7e-116;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325925 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 325984

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325985 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 326044

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326045 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 326104

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326105 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 326164

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326165 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAT 326224

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326225 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 326284

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326285 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 326344

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db     326345 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 326380


RESULT 7
AE015131
LOCUS       AE015131               10370 bp    DNA     linear   BCT 18-OCT-2002
DEFINITION  Shigella flexneri 2a str. 301 section 94 of 412 of the complete
            genome.
ACCESSION   AE015131 AE005674
VERSION     AE015131.1  GI:24051313
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 301
  ORGANISM  Shigella flexneri 2a str. 301
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 10370)
  AUTHORS   Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
            Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
            Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
            Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
            Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
  TITLE     Genome sequence of Shigella flexneri 2a: insights into
            pathogenicity through comparison with genomes of Escherichia coli
            K12 and O157
  JOURNAL   Nucleic Acids Res. 30 (20), 4432-4441 (2002)
   PUBMED   12384590
REFERENCE   2  (bases 1 to 10370)
  AUTHORS   Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B.,
            Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y.,
            Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y.,
            Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
            of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.
            China
FEATURES             Location/Qualifiers
     source          1. .10370
                     /organism="Shigella flexneri 2a str. 301"
                     /mol_type="genomic DNA"
                     /strain="301"
                     /serotype="2a"
                     /db_xref="taxon:198214"
     tRNA            complement(127. .214)
                     /product="tRNA-Ser"
                     /note="anticodon: GGA"
     gene            410. .1387
                     /gene="ycdW"
                     /locus_tag="SF1027"
                     /pseudo
     CDS             410. .1387
                     /gene="ycdW"
                     /locus_tag="SF1027"
                     /note="similar to Escherichia coli K12 putative
                     dehydrogenase gi: 1787270 (326 aa). BLAST with identity of
                     99% in 326 aa. This CDS contains an in-frame stop codon.
                     The sequence has been checked and is believed to be
                     correct."
                     /pseudo
                     /codon_start=1
                     /transl_table=11
     gene            1442. .2179
                     /gene="ycdX"
                     /locus_tag="SF1028"
     CDS             1442. .2179
                     /gene="ycdX"
                     /locus_tag="SF1028"
                     /note="Residues 1 to 245 of 245 are 99 pct identical to
                     residues 1 to 245 of a 245 aa protein from Escherichia
                     coli O157:H7 EDL933 ref: NP_287168.1"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, conserved hypothetical protein"
                     /protein_id="AAN42652.1"
                     /db_xref="GI:24051314"
                     /translation="MYPVDLHMHTVASTHAYSTLSDYIAQAKQKGIKLFAITDHGPDM
                     EDAPHHWHFINMRIWPRVVDGVGILRGIEANIKNVDGEIDCSGKMFDSLDLIIAGFHE
                     PVFAPHDKATNTQAMISTIASGNVHIISHPGNPKYEIDVKAVAEAAAKHQVALEINNS
                     SFLHSRKGSEDNCRAVAAAVRDAGGWVALGSDSHTAFTMGEFEECLKILDAVDFPLER
                     ILNVSPRRLLNFLESRGMAPIAEFADL"
     gene            2203. .2757
                     /gene="ycdY"
                     /locus_tag="SF1029"
     CDS             2203. .2757
                     /gene="ycdY"
```

/locus_tag="SF1029"
/note="Residues 1 to 184 of 184 are 99 pct identical to residues 1 to 184 of a 184 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287169.1"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase component"
/protein_id="AAN42653.1"
/db_xref="GI:24051315"
/translation="MNEFSILCRVLGSLYYRQPQDPLLVPLFTLIREGKLAANWPLEQ DELLTRLQKSCDMTQVSADYNALFIGDECAVPPYRSAWVEDATEAEVRAFLSERGMLL ADTPADHIGTLLLAASWLEDQSTEDESEALETLFSEYLLPWCGAFLGKVEAHATTPFW RTMAPLTRDAISAMWDELEEDSEE"

gene 2811. .3350
/gene="ycdZ"
/locus_tag="SF1030"

CDS 2811. .3350
/gene="ycdZ"
/locus_tag="SF1030"
/note="Residues 1 to 179 of 179 are 98 pct identical to residues 1 to 179 of a 179 aa protein from Escherichia coli K12 ref: NP_415554.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42654.1"
/db_xref="GI:24051316"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIWGWVAVSLGLL SWAGFLGCTAYFACPQGGLKGLAISAATLLSGVVWAMVIIYGSALAPHLEILGYVITG IVAFLMCIQAKQLLLSFVPGTFIGACATFAGQGDWKLVLPSLALGLVFGYAMKNSGLW LAARSAKTAHREQEIKNKA"

gene complement(3414. .4189)
/gene="csgG"
/locus_tag="SF1031"
/pseudo

CDS complement(3414. .4189)
/gene="csgG"
/locus_tag="SF1031"
/note="similar to Escherichia coli K12 curli production assembly-transport component, 2nd curli operon gi: 1787274 (278 aa). BLAST with identity of 97% in 257 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11

gene complement(4454. .5662)
/locus_tag="SF1032"

CDS complement(4454. .5662)
/locus_tag="SF1032"
/note="Residues 1 to 402 of 402 are 100 pct identical to residues 1 to 402 of a 402 aa protein from Plasmid R100 ref: NP_052925.1"
/codon_start=1
/transl_table=11
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translation="MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTLTLTEL GRNLPTKARTKHNIKRIDRLLGNRHLHKERLAVYRWHASPICSGNTMPIVLVDWSDIR EQKRLMVLRASVALHGRSVTLYEKAFPLSEQCSKKAHDQFLADLASILPSNTTPLIVS DAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAENWKPISNLHDMSSSHSKTLGYKR LTKSNPISCQILLYKSRSKGRKNQRSTRTHCHHPSPKIYSASAKEPWILATNLPVEIR TPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALMLQLTCWL AGVHAQKQGWDKHPQANTVRNRNVLSTVRLGMEVLRHSGYTITREDSLVAATLLTQNL FTHGYVLGKL"

gene complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"

CDS complement(5995. .6384)
/gene="csgE"
/locus_tag="SF1033"
/note="Residues 1 to 129 of 129 are 98 pct identical to residues 1 to 129 of a 129 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287173.1"
/codon_start=1
/transl_table=11
/product="curli production assembly-transport component, 2nd curli operon"
/protein_id="AAN42656.1"
/db_xref="GI:24051318"
/translation="MKRYLRWIVAAEFLFAAGNLHAVEVEVPGLLTDHTVSSIGHDFY RAPSDKWESDYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPLKRDSEKTVVFAL IQTEEALNRRQINQALLSTDDLAHDEF"

gene complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"

CDS complement(6389. .7069)
/gene="csgD"
/locus_tag="SF1034"
/note="Residues 11 to 226 of 226 are 100 pct identical to residues 1 to 216 of a 216 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=11
/product="putative 2-component transcriptional regulator for 2nd curli operon"
/protein_id="AAN42657.1"
/db_xref="GI:24051319"
/translation="MRSIKKWGFIMFNEVHSIHGHTLLLITKPSLQATALLQHLKQSL AITGKLHNIQRSLDDISSGSIILLDMMEADKKLIHYWQDTLSRKNNNIKILLLNTPED YPYRDIENWPHINGVFYAMEDQERVVNGLQGVLRGECYFTQKLASYLITHSGNYRYNS TESALLTHREKEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVS WANDNLRR"

gene 7766. .8248
/gene="csgB"
/locus_tag="SF1035"

CDS 7766. .8248
/gene="csgB"
/locus_tag="SF1035"
/note="Residues 10 to 160 of 160 are 99 pct identical to residues 1 to 151 of a 151 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_287175.1"
/codon_start=1
/transl_table=11
/product="minor curlin subunit precursor, similar ro CsgA"
/protein_id="AAN42658.1"
/db_xref="GI:24051320"
/translation="MYDQVQGDNMKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAV NELSKSSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGDYNLAYI DQAGSANDASISQGAYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVTQR

Query Match 98.9%; Score 451.2; DB 1; Length 10370;
Best Local Similarity 99.3%; Pred. No. 5.8e-115;
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
Db 7793 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 7852

Qy 61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
Db 7853 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 7912

Qy 121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
Db 7913 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 7972

Qy 181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
Db 7973 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAATCGGGCA 8032

Qy 241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
Db 8033 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 8092

```
Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8093 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 8152

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       8153 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAGTTGTAGTGCAGAGA 8212

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||||||||| ||||||||||||||||||||
Db       8213 CAGTCGCAAATGGCAATTCGCGTGACACAACGTTAA 8248
```

RESULT 8
AE016759

```
LOCUS       AE016759               306358 bp    DNA     linear   BCT 24-DEC-2002
DEFINITION  Escherichia coli CFT073 section 5 of 18 of the complete genome.
ACCESSION   AE016759 AE014075
VERSION     AE016759.1  GI:26107527
KEYWORDS    .
SOURCE      Escherichia coli CFT073
  ORGANISM  Escherichia coli CFT073
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Extensive Mosiac Structure Revealed by the Complete Genome Sequence
            of Uropathogenic Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
  PUBMED    12471157
REFERENCE   2  (bases 1 to 306358)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .306358
                     /organism="Escherichia coli CFT073"
                     /mol_type="genomic DNA"
                     /strain="CFT073"
                     /db_xref="taxon:199310"
     gene            complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
     CDS             complement(83. .3820)
                     /gene="iroC"
                     /locus_tag="c1253"
                     /note="Residues 27 to 1231 of 1245 are 80.16 pct identical
                     to residues  1 to 1209 of 1218 from GenPept.129 :
                     >emb|CAD05883.1|  (AL627276) putative ABC transporter
                     protein [Salmonella enterica subsp. enterica serovar
                     Typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP binding cassette (ABC) transporter homolog"
                     /protein_id="AAN79727.1"
                     /db_xref="GI:26107528"
                     /translation="MRDRSVLAFLRNDPASGSYFTDGGGLMPANHTPTPAQSWIVRLA
                     RVCWERKKLSVIVVVASVSTILLAALTPLLTRQAVNDALAGNPARLPWLACGLLLIAF
                     FDFIGNYVRRGYAGMLSLWVQHTLRGRVFDSIQKLDGAGQDALRTGQVISRTNSDLQQ
                     VHTLLQMCPVPLAVFTYYIAGIAVMLWMSPAMTLIVVCVLVCLAITALRARRRVFAQT
                     GMASDQLANLTEHIREVLAQISVVKSCVAEMRETHWLDRQSRQIVRVRIGAVISQAMP
                     GATMLALPVLGQIVLLCYGGWSVMHGRIDLGTFVAFASFLAMLTGPTRVLASFLVIAQ
                     RTQASVERVFALIDTRSQMEDGTESINSQVVGLELENMSFDYHHGDRHILSDISFSLR
                     AGETVAVVGASGSGKSTLLMLLARFYDPCSGKIWLNTSEGRQNLRDIRLEALRRRVGI
                     VFEDAFLFAGTVAENIAYGHPQATADDIRRAAAAAGASDFINALPKGFDSLLTERGTN
                     LSGGQRQRIALARALITAPDVLILDDTTSAVDAVTEAEINTALGRYADEGHMLLVIAR
                     RRSTLQLASRVVVLDKGRMVDTGTPAELEARCPAFRALMTGDSDFLATSHNSHNELWP
                     AEPATQDDVTDTGDKGFVARMTRVPENAVQQALAGKGRKVTSLLKPVAWMFVIAALLI
                     ALDSAAGVGVLILLQHGIDSGVAAGDMSTIGLCALLALCLVIVGWCSYSLQTVFAARA
                     AESVQHSVRLRSFGHMLRLGLPWHEKHADSRLTRMTVDVDSLARFLQNGLAGAATSLV
                     TMFAIAATMFWLDPILALTALSAVPVAALATMIYRRLSTPAYAQARLEIGKVNSTLQE
                     KVSGLRVVQSHGQQELEGARLRALSERFRATRVRAQKYLAVYFPFLTFCTEASYAAVL
                     LVGASQVAAGEMTAGVLAAFFLLLGQFYGPVQQLSGIVDAWQQATASGKHIDELLATE
                     GTENLGSSSVLPVTGALHLDEVTFSYPDSHEPALNKLTLTIPEGMVVAVVGRSGAGKS
                     TLIKLIAGLYFPTHGNIRIGVQMLDDASLTEYRRQIGLVDQDVALFSSDIAENIRYSR
                     PSATNEDVEIASQRAGLYEMVCNLPQGFRTPVNNGGADLSAGQRQLIALARAQLANAH
                     ILLLDEATSCLDRTSEERLMSSLTDVVHAGKHSALIVAHRLTTAQRCDLIAVIDKGLL
                     AEYGTHEQLLSAGGLYTRLWHDSVSSTALHRQHNMKEETPG"
     gene            complement(3882. .5045)
                     /gene="iroB"
                     /locus_tag="c1254"
     CDS             complement(3882. .5045)
                     /gene="iroB"
                     /locus_tag="c1254"
                     /note="Escherichia coli O157:H7 ortholog: z1190"
                     /codon_start=1
                     /transl_table=11
                     /product="Putative glucosyltransferase"
                     /protein_id="AAN79728.1"
                     /db_xref="GI:26107529"
                     /translation="MRRLPDLDRQAERDFLMRILFVGPPLYGLLYPVLSLAQAFRVNG
                     HEVLIASGGQFAQKAAEAGLVVFDAAPGLDSEAGYRHHEAQRKKSNIGTQMGNFSFFS
                     EEMADHLVEFAGHWRPDLIIYPPLGVIGPLIAAKYDIPVVMQTVGFGHTPWHIRGVTR
                     SLTDAYRRHNVGATPRDMAWIDVTPPSMSILENDGEPIIPMQYVPYNGGAVWEPWWER
                     RPDRKRLLVSLGTVKPMVDGLDLIAWVMDSASEVDAEIILHISANARSDLRSLPSNVR
                     LVDWIPMGVFLNGADGFIHHGGAGNTLTALHAGIPQIVFGQGADRPVNARVVAERGCG
                     IIPGDVGLSSNMINAFLNNRSLRKASEEVAAEMAAQPCPGEVAKSLITMVQKG"
     gene            5535. .5834
                     /locus_tag="c1255"
     CDS             5535. .5834
                     /locus_tag="c1255"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79729.1"
                     /db_xref="GI:26107530"
                     /translation="MKPTLLQTVENICLQILFVLFLFTGTLIDVEALVSCSSVAKNVQ
                     RTPASSVSSSVYSASVSRTAADIMLLSNNPLFAPGQVSFFGCSLAASGIHVTILE"
     repeat_region   5867. .6577
                     /insertion_seq="IS1541A-like"
     gene            5961. .6470
                     /locus_tag="c1256"
     CDS             5961. .6470
                     /locus_tag="c1256"
                     /note="Residues 18 to 169 of 169 are 94.07 pct identical
                     to residues  1 to 152 of 152 from SwissProt.40 :
                     >sp|Q57334|T200_SALTY Transposase for insertion sequence
                     element IS200"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79730.1"
                     /db_xref="GI:26107531"
                     /translation="MRSGNCKCQTRNQKGVPMGNEKSLAHTRWNCKYHIVFAPKYRRQ
                     VFYREKRRAIGCILRKLCEWKSVRILEAECCADHIHMLVEIPPKMSVSGFMGYLKGKS
                     SLMPYEQFGDLKFKYRNREFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPY
                     PGSPFTGRK"
     repeat_region   complement(6578. .6940)
                     /note="incomplete"
                     /insertion_seq="IS4"
     gene            complement(6598. .6960)
                     /locus_tag="c1257"
     CDS             complement(6598. .6960)
                     /locus_tag="c1257"
                     /function="IS, phage, Tn"
                     /note="Residues 8 to 120 of 120 are 98.23 pct identical to
                     residues 316 to 428 of 442 from MG1655 : b4278"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="Putative conserved protein"
                     /protein_id="AAN79731.1"
                     /db_xref="GI:26107532"
                     /translation="MPRQPDQGYREIKQTMQLSRLTLRSKKPELVEQELWGVLLAYNL
                     VRYQMIKMAEHLKGYWPNQLSFSESCGMVMRMLMTLQGASPGRIPELMRDLASMGQLV
                     KLPTRRETAFPRVVKERP"
     gene            7651. .8013
                     /locus_tag="c1258"
     CDS             7651. .8013
                     /locus_tag="c1258"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79732.1"
                     /db_xref="GI:26107533"
                     /translation="MLGACFYSWGTSFYYGFPWWGAELVLMMWPEACFMLLQSSLYSL
                     LDGEQVLFSFWELNKKQYEGFEFYQIFSGNIVASCSSCSGVSSNSSAHYARCTDFMRH
                     YSLNNYAACQLGRQACLS"
     gene            7835. .8269
                     /locus_tag="c1259"
     CDS             7835. .8269
                     /locus_tag="c1259"
                     /note="Residues 1 to 139 of 144 are 50.71 pct identical to
                     residues  104 to 242 of 258 from GenPept.129 :
                     >gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
                     [Shigella flexneri]"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79733.1"
                     /db_xref="GI:26107534"
                     /translation="MKDLSFIRFFLAILLLLVPPALEFPVIHRHITPGALTLCVITAL
                     IITLLASLGGRLVSVKCISEISFIRRHRIECMMAGFMIYFWTFSLIAGWYKPQFKKEY
                     QTLHYENARYYVLARYDGRLVLSQSYRSGSRKFVIINGGHRL"
     gene            8477. .8553
                     /locus_tag="c5534"
     tRNA            8477. .8553
                     /locus_tag="c5534"
                     /product="tRNA-Arg"
                     /function="RNA; tRNA"
     gene            complement(8832. .9290)
                     /locus_tag="c1260"
     CDS             complement(8832. .9290)
                     /locus_tag="c1260"
                     /function="putative enzyme; Transposases Insertion
                     Sequence Associated"
                     /note="Escherichia coli O157:H7 ortholog: z4324"
                     /codon_start=1
                     /transl_table=11
                     /product="Putative Transposase"
                     /protein_id="AAN79734.1"
                     /db_xref="GI:26107535"
                     /translation="MASAADARRIVSHYERRWLIEEYHKAWKSGGTCVESLRMQTRDN
                     LERMVVIKAFIAVRMLGLRQEGISEETQNDSCKKILTPTEWKLLWVKLEGKQLPSQTP
                     TLKWACLKLGRWHDSKRTGRPGWVVMWDGWFRLQDMVEGYPVMKSLDQEI"
     gene            complement(9302. .9667)
                     /locus_tag="c1261"
     CDS             complement(9302. .9667)
                     /locus_tag="c1261"
                     /note="Residues 1 to 35 of 121 are 97.14 pct identical to
                     residues  59 to 93 of 316 from GenPept.129 :
                     >gb|AAK18437.1|AF348706_126 (AF348706) putative
                     transposase [Shigella flexneri]"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /protein_id="AAN79735.1"
                     /db_xref="GI:26107536"
                     /translation="MAERLGEIQKRVITVCDREADIWHYLHYKVSHGQRFVVRTAQNS

Query Match             98.9%;  Score 451.2;  DB 1;  Length 306358;
Best Local Similarity   99.3%;  Pred. No. 5.5e-115;
Matches  453;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44773 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 44832

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      44833 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTTGCGGTAAATGAATTGAGTAAG 44892

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44893 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 44952

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      44953 CGGCAGGGAGGCTCAAAACTTTTGACGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 45012

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45013 AAGATTGATCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 45072

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45073 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 45132

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45133 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 45192

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db      45193 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 45228
```

```
RESULT 9
AE016981
LOCUS       AE016981               292504 bp    DNA     linear   BCT 22-APR-2003
DEFINITION  Shigella flexneri 2a str. 2457T section 4 of 16 of the complete
            genome.
ACCESSION   AE016981 AE014073
VERSION     AE016981.1  GI:30040616
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 2457T
  ORGANISM  Shigella flexneri 2a str. 2457T
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Complete Genome Sequence and Comparative Genomics of Shigella
            flexneri Serotype 2a Strain 2457T
  JOURNAL   Infect. Immun. 71 (5), 2775-2786 (2003)
   PUBMED   12704152
REFERENCE   2  (bases 1 to 292504)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .292504
                     /organism="Shigella flexneri 2a str. 2457T"
                     /mol_type="genomic DNA"
                     /strain="2457T"
                     /serotype="2a"
                     /db_xref="taxon:198215"
     gene            complement(223. .441)
                     /gene="infA"
                     /locus_tag="S0884"
```

```
CDS             complement(223. .441)
                /gene="infA"
                /locus_tag="S0884"
                /function="factor; Proteins - translation and
                modification"
                /note="residues 1 to 72 of 72 are 100.00 pct identical to
                residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
                /codon_start=1
                /transl_table=11
                /product="protein chain initiation factor IF-1"
                /protein_id="AAP16348.1"
                /db_xref="GI:30040617"
                /translation="MAKEDNIEMQGTVLETLPNTMFRVELENGHVVTAHISGKMRKNY
                IRILTGDKVTVELTPYDLSKGRIVFRSR"
gene            complement(726. .1430)
                /gene="aat"
                /locus_tag="S0885"
CDS             complement(726. .1430)
                /gene="aat"
                /locus_tag="S0885"
                /function="enzyme; Aminoacyl tRNA synthetases, tRNA
                modification"
                /note="residues 1 to 234 of 234 are 100.00 pct identical
                to residues 1 to 234 of 234 from Escherichia coli K-12 :
                B0885"
                /codon_start=1
                /transl_table=11
                /product="leucyl, phenylalanyl-tRNA-protein transferase"
                /protein_id="AAP16349.1"
                /db_xref="GI:30040618"
                /translation="MRLVQLSRHSIAFPSPEGALREPNGLLALGGDLSPARLLMAYQR
                GIFPWFSPGDPILWWSPDPRAVLWPESLHISRSMKRFHKRSPYRVTMNYAFGQVIEGC
                ASDREEGTWITRGVVEAYHRLHELGHAHSIEVWREDELVGGMYGVAQGTLFCGESMFS
                RMENASKTALLVFCEEFIGHGGKLIDCQVLNDHTASLGACEIPRRDYLNYLNQMRLGR
                LPNNFWVPRCLFSPQE"
gene            complement(1472. .3193)
                /gene="cydC"
                /locus_tag="S0886"
CDS             complement(1472. .3193)
                /gene="cydC"
                /locus_tag="S0886"
                /function="transport; Transport of small molecules: Other"
                /note="residues 1 to 573 of 573 are 97.03 pct identical to
                residues 1 to 573 of 573 from Escherichia coli K-12 :
                B0886"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding component of cytochrome-related
                transport"
                /protein_id="AAP16350.1"
                /db_xref="GI:30040619"
                /translation="MRALLPYLALYKRHKWMLSLGIVLAIVTLLASIGLLTLSGWFLS
                ASAVAGVAGLYSFNYMLPAAGVRGAAITRTAGRYFERLVSHDATFRVLQHLRIYTFSK
                LLPLSPAGLARYRQGELLNRVVADVDTLDHLYLRVISPLVGAFVVIMVVTIGLSFLDF
                TLAFTLGSVMLLTLFLMPPLFYRAGKSTGQNLTHLRGQYRQQLTAWLQGQAELTIFGA
                SDRYRTQLENTEIQWLEAQRRQSELTALSQAIMLLIGALAVILMLWMASGGVGGNAQP
                GALIALFVFCALAAFEALAPVTGAFQHLGQVIASAVRISDLTDQKPEVTFPDTQTRVA
                DRVSLTLRDVQFTYPEQSQQALKGISLQVNAGEHIAILGRTGCGKSTLLQLLTRAWDP
                QQGEILLNDSPIASLNEAALRQTISVVPQRVHLFSATLRDNLLLASPGSSDEALSEIL
                RRVGLEKLLEDAGLNSWLGEGGRQLSGGELRRLAIARTLLHDAPLVLLDEPTEGLDAT
                TESQILELLAKMMREKTVLMVTHRLRGLSRFQQIIVMDNGQIIEQGTHAELLARQGRY
                YQFKQGL"
gene            complement(3194. .4960)
                /gene="cydD"
                /locus_tag="S0887"
CDS             complement(3194. .4960)
                /gene="cydD"
                /locus_tag="S0887"
                /function="transport; Not classified"
                /note="residues 1 to 588 of 588 are 92.85 pct identical to
                residues 1 to 588 of 588 from Escherichia coli K-12 :
                B0887"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding component of cytochrome-related
                transport"
                /protein_id="AAP16351.1"
                /db_xref="GI:30040620"
                /translation="MNKSRQKELTRWLKQQSVISQRWLNISRLLGFVSGILIIAQAWF
                MARILQHMIMENIPREALLLPFTLLVLTFVLRAWVVWLRERVGYHAGQHIRFAIRRQV
                LDRLQQAGPAWIQGKPAGSWATLVLEQIDDMHDYYARYLPQMALAVSVPLLIVVAIFP
                SNWAAALILLGTAPLIPLFMALVGMGAADANRRNFLALARLSGHFLDRLRGMETLRIF
                GRGEAEIESIRSASEDFRQRTMEVLRLAFLSSGILEFFTSLSIALVAVYFGFSYLGEL
                DFGHYDTGVTLAAGFLALILAPEFFQPLRDLGTFYHAKAQAVGAADSLKTFMETPLAH
                PQRGEAELALTDPLTIEAEDLFITSPEGKTLAGPLNFTLPAGQRAVLVGRSGSGKSSL
                LNALSGFLSYQGSLRINGIELRDLSPESWRKHLSWVGQNPQLPAAKLRDNVLLARPDA
                SEQELQAALDNAWVSEFLPLLPQGVDTPVGDQAARLSVGQAQRVAVARALLNPCSLLL
                LDEPAASLDAHSEQRVMEALNAASLRQTTLMVTHQLEDLADWDVIWVMQDGRIIEQGR
                YAELSVAGGPFATLLAHRQEEI"
gene            complement(5083. .6048)
                /gene="trxB"
                /locus_tag="S0888"
CDS             complement(5083. .6048)
                /gene="trxB"
                /locus_tag="S0888"
                /function="enzyme; 2-Deoxyribonucleotide metabolism"
                /note="residues 1 to 321 of 321 are 95.95 pct identical to
                residues 1 to 321 of 321 from Escherichia coli K-12 :
                B0888"
                /codon_start=1
                /transl_table=11
                /product="thioredoxin reductase"
                /protein_id="AAP16352.1"
                /db_xref="GI:30040621"
                /translation="MGTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQL
                TTTTEVENWPGDPNDLTGPLLMARMHEHATKFETEIIFDHINKVDLQNRPFRLNGDNG
                EYTCDALIIATGASARYLGLPSEEAFKGRGVSACATCDGFFYRNQKVAVIGGGNTAVE
                EALYLSNIASEVHLIHRRDGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGV
                TGVRLRDTQNSDNIESLDVAGLFVAIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQ
                TSIPGVFAAGDVMDHIYRQAITSAGTGCMAALDAERYLDGLADAK"
gene            6592. .7086
                /gene="lrp"
                /locus_tag="S0889"
CDS             6592. .7086
                /gene="lrp"
                /locus_tag="S0889"
                /function="regulator; Central intermediary metabolism:
                Pool, multipurpose conversions"
                /note="residues 1 to 164 of 164 are 100.00 pct identical
                to residues 1 to 164 of 164 from Escherichia coli K-12 :
                B0889"
                /codon_start=1
                /transl_table=11
                /product="regulator for leucine (or lrp) regulon and
                high-affinity branched-chain amino acid transport system"
                /protein_id="AAP16353.1"
                /db_xref="GI:30040622"
                /translation="MVDSKKRPGKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTP
                CLERVRRLERQGFIQGYTALLNPHYLDASLLVFVEITLNRGAPDVFEQFNTAVQKLEE
                IQECHLVSGDFDYLLKTRVPDMSAYRKLLGETLLRLPGVNDTRTYVVMEEVKQSNRLV
                IKTR"
gene            7221. .11249
                /gene="ftsK"
                /locus_tag="S0890"
CDS             7221. .11249
                /gene="ftsK"
                /locus_tag="S0890"
                /function="phenotype; Cell division"
                /note="residues 1 to 737 of 1342 are 68.24 pct identical
                to residues 1 to 737 of 1329 from Escherichia coli K-12 :
                B0890"
                /codon_start=1
                /transl_table=11
                /product="cell division protein"
                /protein_id="AAP16354.1"
                /db_xref="GI:30040623"
```

/translation="MSQEYTEDKEVTLTKLSSGRRLLEALLILIVLFAVWLMAALLSF
NPSDPSWSQTAWHEPIHNLGGMPGAWLADTLFFIFGVMAYTIPVIIVGGCWFAWRHQS
SDEYIDYFAVSLRIIGVLALILTSCGLAAINADDIWYFASGGVIGSLLSTTLQPLLHS
SGGTIALLCVWAAGLTLFTGWSWVTIAEKLGGWILNILTFASNRTRRDDTWVDEDEYE
DDEEYEDENHGKQHESRRARILRGALARRKRLAEKFINPMGRQTDAALFSGKRMDDDE
EITYTARGVAADPDDVLFSGNRATQPEYDEYDPLLNSAPITEPVAVAAAATTATQSWA
APVEPVTQTPPVASVDVPPSQPTVAWQPVPGPQTGEPVIAPAPEGYPQQSQYAQPAVQ
YNEPLQQPVQPQQPYYAPAAEQPAQQPYYAPAPEQPVAGNAWQAEEQQSTFAPLSTYQ
TEQTYQQPAAQEPLYQQPQPVEQQPVVEPEPVVEETKPARPPLYYFEEVEEKRARERE
QLAAWYQPIPEPVKEPEPIKSSLKAPSVAAVPPVEAAAAVSPLASGVKKATLATGAAA
TVAAPVFSLANSGGPRPQVKEGIGPQLPRPKRIRVPTRRELASYGIKLPSQRAAEEKA
REAQRNQYDSGDQYNDDEIDAMQQDELARQFAQTQQQRYGEQYQHDVPVNAEDADAAA
EAELARQFAQTQQQRYSGEQPAGANPFSLDDFEFSPMKALLDDGPHEPLFTPIVEPVQ
QPQQPVAPQQQYQQPQQPVAPQQQYQQPQQQVAPQPQYQQPQQPVAPQQQYQQPQQPV
APQPQYQQPQQPVAPQPQYQQPQQPVAPQPQDTLLHPLLMRNGDSRPLHKPTTPLPSL
DLLTPPPSEVEPVDTFALEQMARLVEARLADFRIKADVVNYSPGPVITRFELNLAPGV
KAARISNLSRDLARSLSTVAVRVVEVIPGKPYVGLELPNKKRQTVYLREVLDNAKFRD

Query Match 98.6%; Score 449.6; DB 1; Length 292504;
Best Local Similarity 99.1%; Pred. No. 1.5e-114;
Matches 452; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210815 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 210874

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210875 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 210934

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     210935 TCTTCATTTAATCAGGCAGCCATAATTGGTCAGGCTGGGACTAATAATAGTGCTCAGTTA 210994

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     210995 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAATCGGGCA 211054

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211055 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 211114

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211115 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 211174

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     211175 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAGTTGTAGTGCAGAGA 211234

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||||||||| ||||||||||||||||||||
Db     211235 CAGTCGCAAATGGCAATTCGCGTGACACAACGTTAA 211270
```

RESULT 10
CSP515700
LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
ACCESSION AJ515700
VERSION AJ515700.1 GI:31790491
KEYWORDS csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation component of curlin monomers; regulatory protein.
SOURCE Citrobacter sp. Fec2
ORGANISM Citrobacter sp. Fec2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.
REFERENCE 1
AUTHORS Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
JOURNAL Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE 2 (bases 1 to 2889)
AUTHORS Romling,U.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES Location/Qualifiers
source 1. .2889
/organism="Citrobacter sp. Fec2"
/mol_type="genomic DNA"
/isolate="Fec2"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:213763"
/country="Germany"
gene complement(212. .862)
/gene="csgD"
CDS complement(212. .862)
/gene="csgD"
/function="necessary for transcription of the csgBA operon"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"
/translation="MFNEFHSIHGHTLLLITKPSLQATALLQHLKQSLALSGKLHNIQ
RSLDDISSSCIVLVDMMEADKKLIHYWQDNLSRKNNNLKTLLLNTPDDYPYRDIENWP
HINGVFYVADDEERVVNGLQGILRGECYFSQKLASYLITHSGNYRYNSSESALLTHRE
KEILNKLRIGASNIEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
gene 1623. .2078
/gene="csgB"
CDS 1623. .2078
/gene="csgB"
/function="nucleator"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSSFN
QAAIIGQVGTNNSAKMRQEGSKLLSVVSQEGGSNRAKVDQSGAYNFAYIAQSGHSNDA
SISQSNYGNTAMIIQKGSGNKANITQYGTQKTAVVVQRQSQMAIRVIQR"
gene 2119. .2571
/gene="csgA"
CDS 2119. .2571
/gene="csgA"
/function="curli subunit"
/codon_start=1
/transl_table=11
/product="curlin-csgA protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLLQVAAFAAIVVSGSALAGSVPQWGGGGGGGGSSSGPESTL
SIYQSGVNNAALALQSDARKSDTTIHQNGFGNGADVGQGSDNSTIDLTQNGFKNNATI
DQWNGKNSDITVSQYGGHNAALVNQTASDSSVLVHQVGFGNNATANQY"
ORIGIN

Query Match 78.2%; Score 356.8; DB 1; Length 2889;
Best Local Similarity 86.4%; Pred. No. 1.3e-88;
Matches 394; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||  ||
Db       1623 ATGAAAAACAAATTGTTATTTATGATGTTAACAGTACTGGGTGCGCCTGGAATTGCTTCC 1682

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||| || |||||||||||||  |||| |||||||||| |||||||||||||| ||||||
Db       1683 GCAACAAGTTATGATTTAGCGCATTCCGAATATAACTTTGCGGTAAATGAATTAAGTAAG 1742

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||||||||||| |||||||||  || || ||||||||||  || |
Db       1743 TCTTCATTTAATCAGGCAGCCATTATTGGTCAAGTGGGCACGAATAATAGTGCAAAGATG 1802

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
```

```
Db       1803 CGCCAGGAGGGCTCTAAACTTTTGTCGGTAGTTTCACAAGAAGGTGGAAGTAACCGGGCC 1862

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Db       1863 AAAGTTGACCAATCAGGAGCTTATAATTTTGCGTATATTGCTCAGTCGGGTCATTCCAAT 1922

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Db       1923 GATGCCAGTATATCGCAAAGTAATTACGGTAATACTGCGATGATTATCCAGAAAGGTTCT 1982

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Db       1983 GGAAATAAAGCAAATATTACTCAGTATGGTACGCAAAAAACAGCAGTTGTAGTGCAGAGA 2042

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456

Db       2043 CAGTCGCAAATGGCAATTCGCGTTATTCAACGCTAA 2078
```

RESULT 11
SEU43280

```
LOCUS       SEU43280                2067 bp    DNA     linear   BCT 14-FEB-1996
DEFINITION  Salmonella enteritidis agfBAC operon: fimbrin-like protein
            precursor (agfB), thin aggregative fimbriae precursor (agfA), and
            AgfC (agfC) genes, complete cds.
ACCESSION   U43280
VERSION     U43280.1  GI:1184712
KEYWORDS    .
SOURCE      Salmonella enteritidis
  ORGANISM  Salmonella enteritidis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Doran,J.L., Collinson,S.K., Burian,J., Sarlos,G., Todd,E.C.,
            Munro,C.K., Kay,C.M., Banser,P.A., Peterkin,P.I. and Kay,W.W.
  TITLE     DNA-based diagnostic tests for Salmonella species targeting agfA,
            the structural gene for thin, aggregative fimbriae
  JOURNAL   J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
  MEDLINE   94013373
   PUBMED   8104955
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Clouthier,S.C., Doran,J.L., Banser,P.A. and
            Kay,W.W.
  TITLE     Salmonella enteritidis agfBAC operon encoding thin, aggregative
            fimbriae
  JOURNAL   J. Bacteriol. 178 (3), 662-667 (1996)
  MEDLINE   96146512
   PUBMED   8550497
REFERENCE   3  (bases 1 to 2067)
  AUTHORS   Collinson,S.K., Doran,J.L., Banser,P.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
            Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
            Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Salmonella enteritidis"
                     /mol_type="genomic DNA"
                     /strain="27655-3b"
                     /db_xref="taxon:592"
                     /map="between putA and pyrC"
     -35_signal      571. .576
     -10_signal      598. .603
     gene            696. .1151
                     /gene="agfB"
     CDS             696. .1151
                     /gene="agfB"
                     /note="fimbrin-like protein precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfB precursor"
                     /protein_id="AAC43598.1"
                     /db_xref="GI:1184713"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     696. .752
                     /gene="agfB"
     mat_peptide     753. .1148
                     /gene="agfB"
                     /product="AgfB"
                     /note="putative signal sequence of 21 amino acids;
                     fimbrin-like protein"
     gene            1193. .1648
                     /gene="agfA"
     CDS             1193. .1648
                     /gene="agfA"
                     /note="thin aggregative fimbrin subunit precursor; major
                     fimbrial subunit of thin aggregative fimbriae precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfA fimbrin precursor"
                     /protein_id="AAC43599.1"
                     /db_xref="GI:1184714"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     1292. .1354
                     /gene="agfA"
     mat_peptide     1352. .1645
                     /gene="agfA"
                     /product="AgfA fimbrin"
                     /note="thin aggregative fimbrin subunit; major fimbrial
                     subunit of thin aggregative fimbriae"
     stem_loop       1667. .1696
     gene            1710. .2036
                     /gene="agfC"
     CDS             1710. .2036
                     /gene="agfC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="AgfC"
                     /protein_id="AAC43600.1"
                     /db_xref="GI:1184715"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
ORIGIN
```

```
  Query Match           73.0%;  Score 332.8;  DB 1;  Length 2067;
  Best Local Similarity  83.1%;  Pred. No. 6.4e-82;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Db        696 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 755

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Db        756 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 815

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Db        816 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 875

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Db        876 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 935

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Db        936 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 995

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
```

```
                ||||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db        996 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 1055

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
                || ||||| || |||||||| ||||| ||||| || ||||| ||| ||||||||||| |
Db       1056 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 1115

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
                |||||||| |||||||||||||| || ||||| |||
Db       1116 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 1151


RESULT 12
STAJ2301
LOCUS       STAJ2301               5103 bp    DNA     linear   BCT 15-NOV-2000
DEFINITION  Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and &
            csgC genes.
ACCESSION   AJ002301
VERSION     AJ002301.1  GI:2739232
KEYWORDS    csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
            csgG gene.
SOURCE      Salmonella typhimurium
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Romling,U., Bian,Z., Hammar,M., Sierralta,W.D. and Normark,S.
  TITLE     Curli fibers are highly conserved between Salmonella typhimurium
            and Escherichia coli with respect to operon structure and
            regulation
  JOURNAL   J. Bacteriol. 180 (3), 722-731 (1998)
  MEDLINE   98117058
   PUBMED   9457880
REFERENCE   2  (bases 1 to 5103)
  AUTHORS   Romling,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
            Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .5103
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="SR-11"
                     /db_xref="taxon:602"
     CDS             <1. .322
                     /note="ORF179a"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA05311.1"
                     /db_xref="GI:2739233"
                     /db_xref="GOA:O54290"
                     /db_xref="SWISS-PROT:O54290"
                     /translation="CTLLSGMVWALVIIHGSALAPHLEIVSYVLTGIVAFLMCIQAKQ
                     LLLSFVPGTFIGACATFAGQGDWRLVLPSLALGLIFGYAMKNSGLWLASRREQHSANT
                     AVTK"
     gene            complement(361. .1194)
                     /gene="csgG"
     CDS             complement(361. .1194)
                     /gene="csgG"
                     /function="lipoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgG protein"
                     /protein_id="CAA05312.1"
                     /db_xref="GI:2739234"
                     /db_xref="GOA:O54291"
                     /db_xref="SWISS-PROT:O54291"
                     /translation="MPRLLILVAVLLLSGCLTAPPKQAAKPTLMPRAQSYKDLTHLPA
                     PTGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDSRWFIPLERQGLQ
                     NLLNERKIIRAAQENGTVAMNNRIPLQSLTAANIMVEGSIIGYESNVKSGGVGARYFG
                     IGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGE
                     IGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRELSVPPE
                     S"
     gene            complement(1221. .1637)
                     /gene="csgF"
     CDS             complement(1221. .1637)
                     /gene="csgF"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgF protein"
                     /protein_id="CAA05313.1"
                     /db_xref="GI:2739235"
                     /db_xref="SWISS-PROT:O54292"
                     /translation="MRVKHAVVLLMLFSPLTWAGNMTFQFRNPNFGGNPNNGSFLLNS
                     AQAQNSYKDPAYDNDFGIETPSALDNFTQAIQSQILGGLLTNINTGKPGRMVTNDFII
                     DIANRDGQLQLNVTDRKTGRTSTIEVSGLQTQSTDF"
     gene            complement(1664. .2059)
                     /gene="csgE"
     CDS             complement(1664. .2059)
                     /gene="csgE"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgE protein"
                     /protein_id="CAA05314.1"
                     /db_xref="GI:2739236"
                     /db_xref="SWISS-PROT:O54293"
                     /translation="MKRYLTWIVAAELLFATGNLHANEVEVEVPGLLTDHTVSSIGHE
                     FYRAFSDKWESEYTGNLTINERPSARWGSWITITVNQDVIFQTFLFPMKRDFEKTVVF
                     ALAQTEEALNRRQIDQTLLSTSDLARDEF"
     gene            complement(2064. .2714)
                     /gene="csgD"
     CDS             complement(2064. .2714)
                     /gene="csgD"
                     /function="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgD protein"
                     /protein_id="CAA05315.1"
                     /db_xref="GI:2739237"
                     /db_xref="GOA:O54294"
                     /db_xref="SWISS-PROT:O54294"
                     /translation="MFNEVHSSHGHTLLLITKPSLQATALLQHLKQSLAITGKLHNIQ
                     RSLEDISAGCIVLMDMMEADKKLIHYWQDNLSRKNNNIKTLLLNTPDDYPYREIENWP
                     HINGVFYATEDQEHVVSGLQGILRGECYPSQKLASYLITHSGNYRYNSTESALLTHRE
                     KEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR"
     gene            complement(2869. .2896)
                     /gene="csgDEFG"
     -10_signal      complement(2869. .2874)
                     /gene="csgDEFG"
     -35_signal      complement(2891. .2896)
                     /gene="csgDEFG"
     gene            3344. .3376
                     /gene="csgBA"
     -35_signal      3344. .3349
                     /gene="csgBA"
     -10_signal      3371. .3376
                     /gene="csgBA"
     gene            3469. .3924
                     /gene="csgB"
     CDS             3469. .3924
                     /gene="csgB"
                     /function="nucleator"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgB protein"
                     /protein_id="CAA05316.1"
                     /db_xref="GI:2739238"
                     /db_xref="GOA:P55226"
                     /db_xref="SWISS-PROT:P55226"
                     /translation="MKNKLLFMMLTILGAPGIATATNYDLARSEYNFAVNELSKSSFN
                     QAAIIGQVGTDNSARVRQEGSKLLSVISQEGGNNRAKVDQAGNYNFAYIEQTGNANDA
                     SISQSAYGNSAAIIQKGSGNKANITQYGTQKTAVVVQKQSHMAIRVTQR"
     sig_peptide     3469. .3531
```

```
                     /gene="csgB"
     mat_peptide     3532. .3921
                     /gene="csgB"
                     /product="CsgB protein"
                     /function="nucleator"
     gene            3966. .4421
                     /gene="csgA"
     CDS             3966. .4421
                     /gene="csgA"
                     /function="fiber subunit"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CsgA protein"
                     /protein_id="CAA05317.1"
                     /db_xref="GI:2739239"
                     /db_xref="GOA:P55225"
                     /db_xref="SWISS-PROT:P55225"
                     /translation="MKLLKVAAFAAIVVSGSALAGVVPQWGGGGNHNGGGNSSGPDST
                     LSIYQYGSANAALALQSDARKSETTITQSGYGNGADVGQGADNSTIELTQNGFRNNAT
                     IDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFGNNATANQY"
     sig_peptide     3966. .4025
                     /gene="csgA"
     mat_peptide     4026. .4418
                     /gene="csgA"
                     /product="CsgA protein"
                     /function="fiber subunit"
     gene            4483. .4809
                     /gene="csgC"
     CDS             4483. .4809
                     /gene="csgC"
                     /codon_start=1
                     /transl_table=11
                     /product="CsgC protein"
                     /protein_id="CAA05318.1"
                     /db_xref="GI:2739240"
                     /db_xref="SWISS-PROT:P55227"
                     /translation="MHTLLLLAALSNQITFTTTQQGDIYTVIPQVTLNEPCVCQVQIL
                     SVRDGVGGQSHTQQKQTLSLPANQPIELSRLSVNISSEDSVKIIVTVSDGQSLHLSQQ
                     WPPSAQ"
     sig_peptide     4483. .4533
                     /gene="csgC"
     mat_peptide     4534. .4806
                     /gene="csgC"
                     /product="CsgC"
     gene            4940. .5103
                     /gene="ORF103"
     CDS             4940. .>5103
                     /gene="ORF103"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA05319.1"
                     /db_xref="GI:2739241"
                     /db_xref="SWISS-PROT:O54295"
                     /translation="MSVIKKNIPAIGLCICAFFIHSAVGQQTVQGGVIHFRGAIVEPL
                     CDISTHAENI"
ORIGIN

  Query Match           73.0%;  Score 332.8;  DB 1;  Length 5103;
  Best Local Similarity  83.1%;  Pred. No. 6.3e-82;
  Matches  379;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||
Db       3469 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 3528

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  ||||||| | |||  |||||| ||||| || ||||||||||||||| || |||
Db       3529 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 3588

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||| ||||| ||||||||||  || ||  ||||||||||     ||
Db       3589 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 3648

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| |||||||| ||| | ||| || | |||||||| | |  || |||||
Db       3649 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 3708

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | |||||| |||| |||||||| |||| |||||||| ||  |||||| |||||||
Db       3709 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 3768

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | |||||| ||||||| |||| |||||||||||||||||
Db       3769 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 3828

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              || ||||| || |||||||| ||||| |||||| || ||||| ||| ||||||||||| |
Db       3829 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 3888

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db       3889 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 3924


RESULT 13
AE008749
LOCUS       AE008749               22411 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 53 of 220 of the complete
            genome.
ACCESSION   AE008749 AE006468
VERSION     AE008749.1  GI:16419641
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22411)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22411)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
```

```
        as compressions and repeats; all regions were covered by sequence
        from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .22411
                     /organism="Salmonella typhimurium LT2"
                     /mol_type="genomic DNA"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            434. .1308
                     /gene="phoH"
                     /note="synonym: STM1126"
     RBS             434. .439
                     /gene="phoH"
                     /note="putative RBS for phoH; RegulonDB:STMS1H001398"
     CDS             454. .1308
                     /gene="phoH"
                     /note="similar to E. coli PhoB-dependent, ATP-binding pho
                     regulon component; may be helicase; induced by P
                     starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
                     aa), 92% identity in aa 71 - 354"
                     /codon_start=1
                     /transl_table=11
                     /product="PhoB-dependent ATP-binding pho regulon
                     component"
                     /protein_id="AAL20057.1"
                     /db_xref="GI:16419642"
                     /translation="MCAHNRVQENAGDVYLQLKVLSMGRQKAVIKARREAKRVLRRDS
                     RSHKQREEESVTSLVQMGGVEAIGMARDSRDTSPVKARNEAQAHYLNAIDSKQLIFAT
                     GEAGCGKTWISAAKAAEALIHKDVERIIVTRPVLQADEDLGFLPGDIAEKFAPYFRPV
                     YDVLLKRLGASFMQYCLRPEIGKVEIAPFAYMRGRTFENAVVILDEAQNVTAAQMKMF
                     LTRLGENVTVIVNGDITQCDLPRGVRSGLSDALERFEEDEMVGIVHFNKDDCVRSALC
                     QRTLHAYS"
     gene            complement(1414. .2303)
                     /gene="STM1127"
     CDS             complement(1414. .2295)
                     /gene="STM1127"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
                     identity in aa 15 - 264"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /protein_id="AAL20058.1"
                     /db_xref="GI:16419643"
                     /translation="MEPQPPRLKPGKILDTLGAMQKSLTRASQRIAQYILAFPRQVTQ
                     SSIADLSRETQAGEATVIRFCRTLGYKGFQDFKMDLAIELATTESDDSSPLLDAEVSE
                     SDDAHAIGLKLQNTISNVLSETLNLLDMQQVLGVVDALRHCHSVYIFGVGSSGITALD
                     MKHKLMRIGLRGDAVSNNHFMYMQATLLKAGDVAMGVSHSGTSPETVHSLRLARQAGA
                     TTVAITHNLGSPLCEEADFCLINGNRQGMLQGDSIGTKAAQLFVFDLLYTLLVQSSPE
                     QARESKLRTMNALDMTK"
     RBS             complement(2298. .2303)
                     /gene="STM1127"
                     /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
     gene            complement(2580. .4088)
                     /gene="STM1128"
     CDS             complement(2580. .4076)
                     /gene="STM1128"
                     /note="similar to E. coli putative cotransporter
                     (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
                     identity in aa 7 - 478"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sodium/glucose cotransporter"
                     /protein_id="AAL20059.1"
                     /db_xref="GI:16419644"
                     /translation="MITHSFGIVNYLVLFGYLLAMMLVGVYFSRRQKTADDYFRGGGR
                     VPGWAAGVSVFATTLSSITFMSIPAKAFTSDWTFIIGQYLAIAILPLVFYFYIPFFRK
                     LKVTSAYEYLEARFDVRCRLFASMSFMLFHIGRIAIITFLTVLALRPFIAIDPVILVL
                     LISVMCIIYTWMGGIEGVIWTDVIQGLLLSGSAILIFIVICLKVQGGIDEIFTVTQQA
                     DKFPPATQPHWSWTESTVPVLMIGFLFANIQQFTASQDVVQRYIVTDSIEETKKTLLT
                     NAKLVAVIPVFFFAIGSALFVYYQQHPQLLPAGFNTGGILPLFVVTEMPVGIAGLIIA
                     AIFAAAQSSISSSLNSISSCFNSDIYQRLSHKKRTPENRMKIAKLVILVAGLISSAAS
                     VWLVMADESEIWDAFNSLIGLMGGPMTGLFMLGIFFKRANAGSAVLGIIISVITVLGA
                     RYATDLNFFFYGVIGSLSVVISGVIFAPLFAPAPPLTLDEKPEPKVTL"
     RBS             complement(4083. .4088)
                     /gene="STM1128"
                     /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
     gene            complement(4413. .5099)
                     /gene="STM1129"
     CDS             complement(4413. .5093)
                     /gene="STM1129"
                     /note="similar to E. coli putative enzyme (AAC76255.1);
                     Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
                     226"
                     /codon_start=1
                     /transl_table=11
                     /product="putative inner membrane protein"
                     /protein_id="AAL20060.1"
                     /db_xref="GI:16419645"
                     /translation="MSLLARLEQSVHENGGLIVSCQPVPGSPMDKPEIVAAMAQAAAS
                     AGAVAVRIEGIENLRTVRPHLSVPIIGIIKRDLTGSPVRITPYLQDVDALAQAGADII
                     AFDASFRSRPVDIDSLLTRIRLHGLLAMADCSTVNEGISCHQKGIEFIGTTLSGYTGP
                     ITPVEPDLAMVTQLSHAGCRVIAEGRYNTPALAANAIEHGAWAVTVGSAITRIEHICQ
                     WFSHAVKR"
     RBS             complement(5094. .5099)
                     /gene="STM1129"
                     /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
     gene            5599. .6759
                     /gene="STM1130"
     CDS             5599. .6759
                     /gene="STM1130"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
                     identity in aa 40 - 403"
                     /codon_start=1
                     /transl_table=11
                     /product="putative inner membrane protein"
                     /protein_id="AAL20061.1"
                     /db_xref="GI:16419646"
                     /translation="MGMQMKNFKKMMTLMALCFSVAITTSGYATTLPDIPEPLKNGTG
                     AIDNNGVIYVGLGTAGTSWYKIDLKKQHKDWERIKSFPGGAREQSVSVFLNDELYVFG
                     GVGKKNSESPLQVYSDVYKYSPVKNTWQKVDTISPVGLTGHTGVKLNETMVLITGGVN
                     EHIFDKYFIDIAAAAADESEKNKVIYNYFNKPAKDYFFNKIVFIYNAKENTWKNAGEL
                     PGAGTAGSSSVMGNNFLMLINGELKPGLRTDVIYRAMWDNDKLTWLKNSQLPPSPGEQ
                     QQEGLAGAFSGYSHGVLLVGGGANFPGAKQNYTNGKFYSHEGINKKWRDEVYGLINGH
                     WQYMGKMKQPLGYGVSVSYGDEVFLIGGENAKGKPVSSVTSFTMRDGNLLIK"
     gene            6793. .7497
                     /gene="STM1131"
     RBS             6793. .6798
                     /gene="STM1131"
                     /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
     CDS             6805. .7497
                     /gene="STM1131"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
                     identity in aa 4 - 239"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL20062.1"
                     /db_xref="GI:16419647"
                     /translation="MKINKYLLGMVSFIAFSSYLQAATLDYRHEYADRTRINKDRIAI
                     IEKLPNGIGFYVDASVKSGGVDGEQDKHLSDLVANAIELGVSYNYKVTDNFVLQPGFI
                     FESGPDTSIYKPYLRGQYNFDSGVYMAGRYRYDYARKTANYSDDEKTNRFDTYIGYVF
                     DELKLEYNFTWMDSDQIKFDNKKTNYEHNVALAWKLNKSFTPYVEVGNVAVRNNTDER
                     QTRYRVGLQYHF"
     gene            7768. .9060
                     /gene="STM1132"
     RBS             7768. .7773
                     /gene="STM1132"
                     /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
     CDS             7780. .9060
                     /gene="STM1132"
                     /note="similar to E. coli putative transport protein
```

(AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57% identity in aa 25 - 418"
/codon_start=1
/transl_table=11
/product="putative sugar transport protein"

Query Match 73.0%; Score 332.8; DB 1; Length 22411;
Best Local Similarity 83.1%; Pred. No. 6.2e-82;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17272 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 17331

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  |||||||  | |||  ||||||  |||||  ||  ||||||||||||||  || |||
Db      17332 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 17391

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||  |||||  ||||||||||  ||  ||  ||||||||||     ||
Db      17392 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 17451

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||  ||||  |||  ||||||||  |||  |  |||  ||  |  ||||||||    |  |  ||  |||||
Db      17452 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 17511

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  |  ||||||  ||||  ||||||||  ||||  ||||||||  ||  ||||||  |||||||
Db      17512 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 17571

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||| ||||||  |  |||||  |||||||  ||||    ||||||||||||||||||
Db      17572 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 17631

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||  |||||  ||  ||||||||  |||||  |||||  ||  |||||  |||  ||||||||||||  |
Db      17632 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 17691

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||  ||||||||||||||  ||  |||||  |||
Db      17692 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 17727
```

RESULT 14
AL627269
LOCUS AL627269 254050 bp DNA linear BCT 04-JUL-2003
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome; segment 5/20.
ACCESSION AL627269 AL513382
VERSION AL627269.1 GI:16502231
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi
ORGANISM Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 254050)
AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T., Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N., Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P., Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.
TITLE Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18
JOURNAL Nature 413 (6858), 848-852 (2001)
MEDLINE 21534947
PUBMED 11677608
REFERENCE 2 (bases 1 to 254050)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).

FEATURES Location/Qualifiers

source 1. .254050
/organism="Salmonella enterica subsp. enterica serovar Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"

gene 181. .2793
/gene="STY1078"
/note="synonym: pepN"

CDS 181. .2793
/gene="STY1078"
/note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
/codon_start=1
/transl_table=11
/product="aminopeptidase N"
/protein_id="CAD08183.1"
/db_xref="GI:16502232"
/db_xref="GOA:Q8Z7T0"
/db_xref="SPTREMBL:Q8Z7T0"
/translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR
HGAPDAPLRLDGEDLTLVSIHVNDAPWTAYKEEEGALIISDLPERFTLRIVNEISPAA
NTALEGLYQSGDALCTQCEAEGFRHITWYLDRPDVLARFTTKIIADKSKYPFLLSNGN
RVAQGELENGRHWVQWQDPFPKPCYLFALVAGDFDVLRDTFTTRSGREVALELYVDRG
NLDRAPWAMTSLKNSMKWDETRFGLEYDLDIYMIVAVDFFNMGAMENKGLNIFNSKYV
LARTDTATDKDYLDIERVIGHEYFHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDL
GSRAVNRISNVRTMRGLQFAEDASPMAHPIRPDKVIEMNNFYTLTVYEKGAEVIRMIH
TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDASNVDLSHFRRWYSQSGTPIVTV
KDDYNPETEQYTLTISQRTPATADQAEKQPLHIPFAIELYDNEGNVIPLQKGGHPVNA
VLNVTQAEQTFTFDNVYFQPVPALLCEFSAPVKLEYKWSDQQLTFLMRHARNDFSRWD
AAQSLLATYIKLNVARHQQGQPLSLPVHVADAFRAVLLDEKIDPALAAEILTLPSANE
IAELFEVIDPIAIAQVREALTRTLAAELADEFLAIYNANHLDEYRVDHGDIGKRTLRN
ACLRFLTFGETELANTLVSKQYRDANNMTDALAALSAAVAAQLPCRDTLMQEYDDKWH
QDGLVMDKWFILQSTSPAENVLETVRSLLKHRSFSMSNPNRIRSLIGAFAGSNPAAFH
AQDGSGYQFLVEMLTDLNSRNPQVASRLIEPLIRLKRYDDKRQEKMRAALEQLKGLEN
LSGDLYEKITKALA"

misc_feature 532. .1332
/gene="STY1078"
/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase family M1, score 245.10, E-value 1.9e-84"

misc_feature 1060. .1089
/gene="STY1078"
/note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature"

gene 3001. .4011
/gene="STY1079"
/note="synonym: pyrD"

CDS 3001. .4011
/gene="STY1079"
/note="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q8Z7S9"
/db_xref="SWISS-PROT:Q8Z7S9"
/translation="MYYPFVRKALFQLDPERAHEFTFQQLRRITGTPLEALVRQKVPT
KPVTCMGLTFKNPLGLAAGLDKDGECIDALGAMGFGSLEIGTVTPRPQPGNDKPRLFR
LVDAEGLINRMGFNNLGVDNLVENVKKAHFDGILGINIGKNKDTPVENGKDDYLICME
KVYAYAGYIAINISSPNTPGLRTLQYGDALDDLLTAIKNKQNDLQVIHHKYVPVAVKI
APDLCEEELIQVADSLLRHNIDGVIATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKST
EIIRRLSLELKGQLPIIGVGGIDSVIAAREKIAAGATLVQIYSGFIFKGPPLIKEIVT
HI"

```
misc_feature    3136. .3990
                /gene="STY1079"
                /note="Pfam match to entry PF01180 DHOdehase,
                Dihydroorotate dehydrogenase, score 563.50, E-value
                1.4e-165"
misc_feature    3238. .3297
                /gene="STY1079"
                /note="PS00911 Dihydroorotate dehydrogenase signature 1"
misc_feature    3874. .3936
                /gene="STY1079"
                /note="PS00912 Dihydroorotate dehydrogenase signature 2"
gene            4174. .4719
                /gene="STY1080"
CDS             4174. .4719
                /gene="STY1080"
                /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
                YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
                discrepency in position of translational start site."
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAD08185.1"
                /db_xref="GI:16502234"
                /db_xref="SPTREMBL:Q8Z7S8"
                /translation="MMRIKPDDNWRWYYDEEHDRMMLDLANGMLFRSRFSRKMLTPDA
                FCPTGFCVDDAALYFSFEEKCRDFELTKEQRAELVLNALVAIRYLKPQMPKSWHFVAH
                GEMWTPGTGDAASVWLSDTAEQVNLLVVEPGENAALCLLAQPGVVIAGRTMQLGDAIK
                IMNDRLKPQVHCHSFSLEQAV"
gene            complement(4716. .5825)
                /gene="STY1081"
CDS             complement(4716. .5825)
                /gene="STY1081"
                /note="Orthologue of E. coli P75863; Fasta hit to P75863
                (369 aa), 86% identity in 367 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAD08186.1"
                /db_xref="GI:16502235"
                /db_xref="GOA:Q8Z7S7"
                /db_xref="SPTREMBL:Q8Z7S7"
                /translation="MATLSRLFIHPVKSMRGIGLTHALADISGLAFDRIFMVTESDGT
                FITARQFPQMVRFTPSPLHDGLHLTAPDGSSALVRFTDFTPQDAPTEVWGNHFTARVA
                PTAINQWLSGFFSRDVQLCWVGPQLTRRVKRHNAVPLGFADGYPYLLTNEASLRDLQQ
                RCPAGVQMEQFRPNLVVSGVAAWEEDSWKVLRIGDVIFDVVKPCSRCIFTTVSPEKGQ
                KHPSGEPLATLQAFRTAQDNGDVDFGQNLIARNSGVIRVGDEVEILATAPAKAYGATT
                LDDSVTPEKHPDGSVTIDWQGQTFCGNNQQVLLEQLENQGIRIPYSCRAGICGCCRIR
                LLEGEVSPLKKSAMGDDGTILSCSCVPKTALRLEN"
misc_feature    complement(4737. .4955)
                /gene="STY1081"
                /note="Pfam match to entry PF00111 fer2, 2Fe-2S
                iron-sulfur cluster binding domains, score 55.90, E-value
                8.5e-13"
gene            5924. .8032
                /gene="STY1082"
CDS             5924. .8032
                /gene="STY1082"
                /note="Similar to Haemophilus influenzae hypothetical
                protein hi0116 HI0116 SW:YCBY_HAEIN (P44524; P43945) fasta
                scores: E(): 0, 60.3% id in 710 aa
                Orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
                (702 aa), 92% identity in 702 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAD08187.1"
                /db_xref="GI:16502236"
                /db_xref="GOA:Q8Z7S6"
                /db_xref="SPTREMBL:Q8Z7S6"
                /translation="MNSLFASTARGLEELLKIELEKLGAVGCQVVQGGVHFQGDTRLI
                YQSLMWSRLASRIILPMGECKVYSDLDLYFGVQAINWTEIFNPGATFAVHFSGLNDTI
                RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRESPDLRINVWLNKETASIALDLSGDGLH
                LRGYRDRTGLAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMWATDRAPGL
                HRGHWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIERARSNAR
                RAGIGELITFEVKDVAQLSNPLPKGPYGTVISNPPYGERLDSDPALIALHSLLGRTMK
                NQFGGWNLSLFSASPDLLGSLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
                EDYANRLRKNLKKLEKWARQEGIECYRLYDADLPEYNVAVDRYGDWAVIQEYAPPKTV
                DAQKARQRLFDIIAATLSVLGIPPNKLVLKTRERQKGKNQYQKMSEKGEFLEVSEYNA
                RLWVNLTDYLDTGLFLDHRIARRMLGEMSKGKDFLNLFSYTGSASVHAGLGGARNTTT
                VDMSRTYLEWAERNLRLNGLSGRAHRLIQADCLGWLREANEQFDLIFIDPPTFSNSKR
                MEESFDVQRDHVALMKDLKRLLRKGGTIMFSNNKRGFRMDLEGLAELGLTAQEITQKT
                LSPDFARNRQIHNCWLIRAA"
misc_feature    5927. .7060
                /gene="STY1082"
                /note="Pfam match to entry PF01170 UPF0020,
                Uncharacterized protein family UPF0020, score 781.30,
                E-value 3.7e-231"
misc_feature    6506. .6541
                /gene="STY1082"
                /note="PS01261 Uncharacterized protein family UPF0020
                signature"
gene            8045. .9952
                /gene="STY1083"
CDS             8045. .9952
                /gene="STY1083"
                /note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
                526 aa overlap
                Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
                overlap
```

Query Match 72.6%; Score 331.2; DB 1; Length 254050;
Best Local Similarity 82.9%; Pred. No. 1.7e-81;
Matches 378; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db      88409 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 88468

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  ||||||| | |||  |||||| |||| || |||||||||||||| || |||
Db      88469 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 88528

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||| |||||  |||||||||||  || ||  ||||||||||     ||
Db      88529 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 88588

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||    |  || |||||
Db      88589 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGAAAATAATCGGGCG 88648

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | ||||||  ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db      88649 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 88708

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||||||| |||||| | ||||| ||||||| ||||   ||||||||||||||||||
Db      88709 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 88768

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| ||||||||||| |
Db      88769 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 88828

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||| |||||||||||||| || ||||| |||
Db      88829 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 88864
```

```
RESULT 15
AE016840/c
LOCUS       AE016840              301983 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
            16 of the complete genome.
ACCESSION   AE016840 AE014613
VERSION     AE016840.1  GI:29137797
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
```

```
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 301983)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .301983
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
     CDS             complement(164. .622)
                     /gene="tnpA"
                     /locus_tag="t1735"
                     /note="corresponds to STY1225 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase for insertion sequence element
                     IS200"
                     /protein_id="AAO69359.1"
                     /db_xref="GI:29137798"
                     /translation="MGDEKSLAHTRWNCKYHIVFAPKYRRQAFYGEKRRAVGSILRKL
                     CEWKNVRILEAECCADHIHMLLEIPPKMSVSSFMGYLKGKSSLMLYEQFGDLKFKYRN
                     REFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSPFTGRK"
     gene            complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
     CDS             complement(844. .1797)
                     /gene="flgL"
                     /locus_tag="t1736"
                     /note="corresponds to STY1223 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 3"
                     /protein_id="AAO69360.1"
                     /db_xref="GI:29137799"
                     /translation="MRISTQMMYEQNMSGITNSQAEWMKLGEQMSTGKRVTNPSDDPI
                     AASQAVVLFQAQAQNSQYALARTFATQKVSLEESVLSQVTTAIQTAQEKIVYAGNGTL
                     SDDDRASLATDLQGIRDQLMNLANSTDGNGRYIFAGYKTEAAPFDQATGGYHGGEKSV
                     TQQVDSARTMVIGHTGAQIFNSITSNAVPEPDGSDSEKNLFVMLDTAIAALKTPVEGN
                     DVEKEKAAAAIDKTNRGLKNSLNNVLTVRAELGTQLSELSTLDSLGSDRALGQKLQMS
                     NLVDVDWNSVISSYVMQQAALQASYKTFTDMQGMSLFQLNR"
     gene            complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
     CDS             complement(1812. .3473)
                     /gene="flgK"
                     /locus_tag="t1737"
                     /note="corresponds to STY1222 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook-associated protein 1"
                     /protein_id="AAO69361.1"
                     /db_xref="GI:29137800"
                     /translation="MSSLINHAMSGLNAAQAALNTVSNNINNYNVAGYTRQTTILAQA
                     NSTLGAGGWIGNGVYVSGVQREYDAFITNQLRGAQNQSSGLTTRYEQMSKIDNLLADK
                     SSSLSGSLQSFFTSLQTLVSNAEDPAARQALIGKAEGLVNQFKTTDQYLRDQDKQVNI
                     AIGSSVAQINNYAKQIANLNDQISRMTGVGAGASPNDLLDQRDQLVSELNKIVGVEVS
                     VQDGGTYNLTMANGYTLVQGSTARQLAAVPSSADPTRTTVAYVDEAAGNIEIPEKLLN
                     TGSLGGLLTFRSQDLDQTRNTLGQLALAFADAFNAQHTKGYDADGNKGKDFFSIGSPV
                     VYSNSNNADKTVSLTAKVVDSTKVQATDYKIVFDGTDWQVTRTADNTTFTATKDADGK
                     LEIDGLKVTVGTGAQKNDSFLLKPVSNAIVDMNVKVTNEAEIAMASESKLDPDVDTGD
                     SDNRNGQALLDLQNSNVVGGNKTFNDAYATLVSDVGNKTSTLKTSSTTQANVVKQLYK
                     QQQSVSGVNLDEEYGNLQRYQQYYLANAQVLQTANALFDALLNIR"
     gene            complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
     CDS             complement(3538. .4488)
                     /gene="flgJ"
                     /locus_tag="t1738"
                     /note="corresponds to STY1221 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar protein FlgJ"
                     /protein_id="AAO69362.1"
                     /db_xref="GI:29137801"
                     /translation="MIGDGKLLASAAWDAQSLNELKAKAGQDPAANIRPVARQVEGMF
                     VQMMLKSMREALPKDGLFSSDQTRLYTSMYDQQIAQQMTAGKGLGLADMMVKQMTGGQ
                     TMPADDAPQVPLKFSLETVNSYQNQALTQLVRKAIPKTPDSSDAPLSGDSKDFLARLS
                     LPARLASEQSGVPHHLILAQAALESGWGQRQILRENGEPSYNVFGVKATASWKGPVTE
                     ITTTEYENGEAKKVKAKFRVYSSYLEALSDYVALLTRNPRYAAVTTAATAEQGAVALQ
                     NAGYATDPNYARKLASMIQQLKAMSEKVSKTYSANLDNLF"
     gene            complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
     CDS             complement(4488. .5591)
                     /gene="flgI"
                     /locus_tag="t1739"
                     /note="corresponds to STY1220 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar P-ring protein precursor"
                     /protein_id="AAO69363.1"
                     /db_xref="GI:29137802"
                     /translation="MYVFKALAGIVLALVATLAHAERIRDLTSVQGVRENSLIGYGLV
                     VGLDGTGDQTTQTPFTTQTLNNMLSQLGITVPTGTNMQLKNVAAVMVTASYPPFARQG
                     QTIDVVVSSMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGGAGASAGGSSVQVN
                     QLNGGRITNGAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRARGYGSATAL
                     DARTVQVRVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCA
                     VAQGNLSVTVNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQSGGSLQSVRSSANLNSVV
                     RALNALGATPMDLMSILQSMQSAGCLRAKLEII"
     gene            complement(5597. .6295)
                     /gene="flgH"
                     /locus_tag="t1740"
     CDS             complement(5597. .6295)
                     /gene="flgH"
                     /locus_tag="t1740"
                     /note="corresponds to STY1219 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar L-ring protein precursor"
                     /protein_id="AAO69364.1"
                     /db_xref="GI:29137803"
                     /translation="MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPV
                     PVANGSIFQSAQPINYGYQPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKT
                     SFGFDTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGN
                     LHVVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQN
                     MGWLQRFFLNLSPM"
     gene            complement(6350. .7132)
                     /gene="flgG"
                     /locus_tag="t1741"
     CDS             complement(6350. .7132)
```

```
                     /gene="flgG"
                     /locus_tag="t1741"
                     /note="distal rod protein; corresponds to STY1218 from
                     Accession AL513382: Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar basal-body rod protein FlgG"
                     /protein_id="AAO69365.1"
                     /db_xref="GI:29137804"
                     /translation="MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDL
                     LYQTIRQPGAQSSEQTTLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGF
                     FQVMLPDGTSAYTRDGSFQVDQNGQLVTAGGFQVQPAITIPANALSITIGRDGVVSVT
                     QQGQAAPVQVGQLNLTTFMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGY
                     VETSNVNVAEELVNMIQVQRAYEINSKAVSTTDQMLQKLTQL"
     gene            complement(7146. .7901)
                     /gene="flgF"
                     /locus_tag="t1742"
     CDS             complement(7146. .7901)
                     /gene="flgF"
                     /locus_tag="t1742"
                     /note="corresponds to STY1217 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="putative flagellar basal-body rod protein FlgF"
                     /protein_id="AAO69366.1"
                     /db_xref="GI:29137805"
                     /translation="MDHAIYTAMGAASQTLNQQAVTASNLANASTPGFRAQLNALRAV
                     PVDGLSLATRTLVTASTPGADMTQGQLDYTSRPLDVALQQDGWLVVQAADGAEGYTRN
                     GNIQVGPTGQLTIQGHPVIGEGGPITVPEGSEITIAADGTISALNPGDPPNTVAPVGR
                     LKLVKAEGNEVQRSDDGLFRLTAEAQAERGAVLAADPSIRIMSGVLEGSNVKPVEAMT
                     DMIANARRFEMQMKVITSVDENEGRANQLLSMS"
     gene            complement(7922. .9133)
                     /gene="flgE"
                     /locus_tag="t1743"
     CDS             complement(7922. .9133)
                     /gene="flgE"
                     /locus_tag="t1743"
                     /note="corresponds to STY1216 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellar hook protein FlgE"
                     /protein_id="AAO69367.1"
                     /db_xref="GI:29137806"
                     /translation="MSFSQAVSGLNAAATNLDVIGNNIANSATYGFKSGTASFADMFA
                     GSKVGLGVKVAGITQDFTDGTTTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKL
```

Query Match 72.6%; Score 331.2; DB 1; Length 301983;
Best Local Similarity 82.9%; Pred. No. 1.7e-81;
Matches 378; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db      37807 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 37748

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  ||||||| | |||  |||||| ||||| || |||||||||||||| || |||
Db      37747 GCGACAAATTATGATCTGGCTCGTTCAGAGTATAATTTTGCGGTAAATGAATTAAGCAAG 37688

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||| ||||| |||||||||| || ||  |||||||||     ||
Db      37687 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 37628

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||     |  || |||||
Db      37627 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGAAAATAATCGGGCG 37568

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | ||||||  ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db      37567 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 37508

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              |||||||||||| |||||| | ||||| ||||||| ||||   ||||||||||||||||||
Db      37507 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCGGCTATTATCCAGAAAGGTTCT 37448

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              || ||||| || |||||||| ||||| ||||| || ||||| ||| |||||||||||| |
Db      37447 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 37388

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||| |||||||||||||| || ||||| |||
Db      37387 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 37352
```

Search completed: March 18, 2004, 15:02:36
Job time : 2314.32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 288.405 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	456	100.0	456	3	AAC64620	Aac64620 Escherich
	2	332.8	73.0	456	3	AAC64618	Aac64618 Salmonell
	3	100	21.9	100	7	ACD68810	Acd68810 E. coli K
	4	100	21.9	100	7	ACD68812	Acd68812 E. coli K
	5	100	21.9	100	7	ACD68811	Acd68811 E. coli K
	6	45.8	10.0	779	1	AAN40272	Aan40272 Sequence
	7	38.2	8.4	4860	8	ADA31934	Ada31934 DNA encod
c	8	37.8	8.3	2000	7	ADA71938	Ada71938 Rice gene
	9	37.4	8.2	456	3	AAC64623	Aac64623 AgfA::PT3
	10	36.6	8.0	456	3	AAC64629	Aac64629 AgfA::PT3
c	11	35.2	7.7	4890	2	AAX34648	Aax34648 Promoter
	12	35	7.7	646	2	AAQ62647	Aaq62647 Fibronect
	13	34.6	7.6	361	2	AAQ73066	Aaq73066 AgfA sequ
	14	34.6	7.6	361	2	AAT74141	Aat74141 Salmonell
	15	34.6	7.6	456	2	AAQ87467	Aaq87467 AgfA sequ
	16	34.6	7.6	456	2	AAT74142	Aat74142 Salmonell
	17	34.6	7.6	456	3	AAC64622	Aac64622 AgfA::PT3
	18	34.6	7.6	456	3	AAC64617	Aac64617 Salmonell
	19	34.6	7.6	456	3	AAC64625	Aac64625 AgfA::PT3
	20	34.6	7.6	456	3	AAC64624	Aac64624 AgfA::PT3
c	21	34.4	7.5	4455	7	ABT17054	Abt17054 Human MP2
c	22	34.4	7.5	5184	9	ADE25651	Ade25651 Human cDN
c	23	34.4	7.5	6325	4	AAK80005	Aak80005 Human imm
c	24	34.4	7.5	9555	7	ABX34542	Abx34542 Human mdd
	25	34.2	7.5	1083	5	AAS76745	Aas76745 DNA encod
	26	34	7.5	12426	4	AAS46482	Aas46482 Tumour su
	27	33.8	7.4	1149	3	AAA65947	Aaa65947 E. coli p
	28	33.8	7.4	1149	7	ACA18571	Aca18571 Prokaryot
	29	33.8	7.4	2655	6	ABN70427	Abn70427 Streptoco
	30	33.8	7.4	2655	6	ABN67963	Abn67963 Streptoco
c	31	33.8	7.4	6108	4	ABL08478	Abl08478 Drosophil
c	32	33.8	7.4	6808	4	ABL08488	Abl08488 Drosophil
	33	33.8	7.4	110000	6	ABN71527_06	Continuation (7 of
	34	33.6	7.4	606	3	AAC43625	Aac43625 Arabidops
	35	33.6	7.4	1299	7	ACA48239	Aca48239 Prokaryot
	36	33.2	7.3	1371	7	ACA29389	Aca29389 Prokaryot
	37	33.2	7.3	4590	5	AAH24065	Aah24065 Yeast AOD
	38	33	7.2	1269	7	ACA48121	Aca48121 Prokaryot
	39	33	7.2	1837	7	ACA47731	Aca47731 Prokaryot
	40	33	7.2	110000	6	ABQ69245_18	Continuation (19 o
c	41	33	7.2	319630	6	ABQ67194	Abq67194 Listeria
	42	32.6	7.1	351	7	ABX54413	Abx54413 Bovine ES
	43	32.6	7.1	967	5	AAS74657	Aas74657 DNA encod
c	44	32.6	7.1	1476	6	AAS97253	Aas97253 Neisseria
c	45	32.6	7.1	1476	6	AAS97288	Aas97288 Neisseria

ALIGNMENTS

RESULT 1
AAC64620
ID AAC64620 standard; DNA; 456 BP.
XX
AC AAC64620;
XX
DT 26-FEB-2001 (first entry)
XX
DE Escherichia coli CsgB DNA sequence SEQ ID NO:4.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Escherichia coli.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36344.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 146 A; 78 C; 113 G; 119 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 3.3e-130;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
```

RESULT 2
AAC64618
ID AAC64618 standard; DNA; 456 BP.
XX
AC AAC64618;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis AgfB DNA sequence SEQ ID NO:2.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36342.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 149 A; 87 C; 109 G; 111 T; 0 U; 0 Other;

Query Match 73.0%; Score 332.8; DB 3; Length 456;
Best Local Similarity 83.1%; Pred. No. 3.1e-92;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  |||||||| | |||  |||||||||||| || ||||||||||||| || |||
Db         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              |||||||||||||||| ||||| ||||||||||  ||  ||  ||||||||||     ||
Db        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| ||| |||||||| ||| | ||| || | ||||||||  | |  || |||||
Db        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  | ||||||  ||||  |||||||| |||| |||||||| ||  |||||| |||||||
Db        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
```

```
                ||||||||||| |||||| | ||||| ||||||| |||    ||||||||||||||||||
Db        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
                || ||||| || |||||||| ||||| ||||| || ||||| ||| ||||||||||| |
Db        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
                |||||||| |||||||||||||| || ||||| |||
Db        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456
```

RESULT 3
ACD68810
ID ACD68810 standard; DNA; 100 BP.
XX
AC ACD68810;
XX
DT 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 80.
XX
KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EP1260592-A1.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MWGB-) MWG-BIOTECH AG.
XX
PI Donner H, Drescher B, Huber A, Weber J;
XX
DR WPI; 2003-241155/24.
XX
PT Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX
PS Claim 3; Page 23; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SQ Sequence 100 BP; 27 A; 19 C; 26 G; 28 T; 0 U; 0 Other;

```
  Query Match             21.9%;  Score 100;  DB 7;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-21;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        266 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 60

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 365
              |||||||||||||||||||||||||||||||||||||||
Db         61 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 100
```

RESULT 4
ACD68812
ID ACD68812 standard; DNA; 100 BP.
XX
AC ACD68812;
XX
DT 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 82.
XX
KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EP1260592-A1.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MWGB-) MWG-BIOTECH AG.
XX
PI Donner H, Drescher B, Huber A, Weber J;
XX
DR WPI; 2003-241155/24.
XX
PT Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX
PS Claim 3; Page 23; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SQ Sequence 100 BP; 28 A; 18 C; 27 G; 27 T; 0 U; 0 Other;

```
  Query Match             21.9%;  Score 100;  DB 7;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-21;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        123 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 182
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 60

Qy        183 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 222
                 |||||||||||||||||||||||||||||||||||||||
Db         61 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 100


RESULT 5
ACD68811
ID   ACD68811 standard; DNA; 100 BP.
XX
AC   ACD68811;
XX
DT   18-SEP-2003  (first entry)
XX
DE   E. coli K12 MG1655 biochip probe SEQ ID 81.
XX
KW   Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS   Escherichia coli.
XX
PN   EP1260592-A1.
XX
PD   27-NOV-2002.
XX
PF   17-MAY-2001; 2001EP-00112179.
XX
PR   17-MAY-2001; 2001EP-00112179.
XX
PA   (MWGB-) MWG-BIOTECH AG.
XX
PI   Donner H,  Drescher B,  Huber A,  Weber J;
XX
DR   WPI; 2003-241155/24.
XX
PT   Biochip containing probes complementary with open reading frames in
PT   Escherichia coli K12, useful for detecting gene expression and expression
PT   patterns.
XX
PS   Claim 3; Page 23; 2004pp; German.
XX
CC   This invention describes a novel biochip comprising probe spots, each
CC   containing many identical probes. The probes are nucleotide sequences of
CC   30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC   least one includes a segment of at least 20 bases identical with, or
CC   complementary to, a segment of an open reading frame (orf) of Escherichia
CC   coli K12. The biochip is used for specific detection of gene expression
CC   in K12 and for determining the gene expression pattern, e.g. for
CC   diagnostic determination of which E. coli strains are present in the gut,
CC   and to determine the effects of e.g. growth media on gene expression. The
CC   biochip provides as comprehensive as possible detection of the K12
CC   genome, with simultaneous analysis of many different genes with a single
CC   device, and comparison of gene expression between K12 and its mutants or
CC   other E. coli strains in a single experiment. Apart from qualitative and
CC   quantitative information about gene expression, it also allows
CC   measurements of population densities for the various strains. The use of
CC   synthetic oligonucleotides for preparation of probes allows free
CC   variation in probe length and ensures high purity (and thus selectivity,
CC   reactivity and reproducibility); also synthetic probes are generally
CC   shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC   ACD81540 represent oligonucleotide probes used with the biochip described
CC   in the invention
XX
SQ   Sequence 100 BP; 31 A; 20 C; 29 G; 20 T; 0 U; 0 Other;

  Query Match             21.9%;  Score 100;  DB 7;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-21;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        208 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 267
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 60
```

```
Qy        268 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 307
                 |||||||||||||||||||||||||||||||||||||||
Db         61 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 100


RESULT 6
AAN40272
ID   AAN40272 standard; cDNA; 779 BP.
XX
AC   AAN40272;
XX
DT   09-JAN-1992  (first entry)
XX
DE   Sequence of a clone numbered as Ag16 which corresponds to the S antigen
DE   of isolate FC27.
XX
KW   Malaria; vaccine; antigen; epitope; immune response; ss.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .777
FT                   /*tag=  d
FT   misc_feature    1. .8
FT                   /*tag=  a
FT                   /label= linker
FT   repeat_region   10. .768
FT                   /*tag=  c
FT                   /rpt_unit= 10. .42
FT                   /note= "final base of certain codons may differ but AA SQ
FT                   encoded stays the same"
FT   misc_feature    772. .779
FT                   /*tag=  b
FT                   /label= linker
XX
PN   WO8402917-A.
XX
PD   02-AUG-1984.
XX
PF   27-JAN-1984;   84WO-AU000016.
XX
PR   28-JAN-1983;   83AU-00007843.
PR   10-JUN-1983;   83AU-00009788.
PR   26-JAN-1984;   84AU-00023842.
XX
PA   (HALL-) HALL INST MED RES.
PA   (KEMP/) KEMP D J.
PA   (AUME-) AUST INST MED RES.
PA   (SARA-) SARAMANE PTY LTD.
XX
PI   Kemp DJ,  Anders RF,  Coppel RL,  Brown G,  Saint R,  Cowman AF;
PI   Mitchell GF;
XX
DR   WPI; 1984-201410/32.
DR   P-PSDB; AAP40376.
XX
PT   Expression of plasmodium falciparum poly:peptide(s) from CDNA - for use
PT   in immunisation against malaria infection.
XX
PS   Claim 5; Fig 1; 62pp; English.
XX
CC   The inventors claim a DNA SQ which comprises a nucleotide sequence
CC   corresp. to all or part of Plasmodium falciparum RNA. Also claimed is a
CC   synthetic peptide or polypeptide displaying the antigenicity of all or
CC   part of a P.falciparum antigen. The base sequence of clone Ag16 (see
CC   AAN40272) indicates that the antigen coded for by this partial
CC   polynucleotide sequence has a homologous repeat structure of 11 AAs
CC   tandemly repeated 23 times
XX
SQ   Sequence 779 BP; 294 A; 125 C; 217 G; 143 T; 0 U; 0 Other;

  Query Match             10.0%;  Score 45.8;  DB 1;  Length 779;
```

Best Local Similarity 47.7%; Pred. No. 0.0011;
Matches 134; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```
Qy        126 ATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCA 185
              ||| |  |  | || | |  | |||||| |||     |  ||  |||  ||  | | ||
Db        198 ATTAGAAGATCCTGCAAAAGCTAGTCAAGGTGGATTAGAAGATCCTGCAAAGGCAAGTCA 257

Qy        186 GGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGAT 245
              ||||| | | ||  |   ||    | |   ||||  ||         | |  |||||
Db        258 AGGAGGATTAGAAGATCCTGCAAAAGCTAGTCAAGGTGGATTAGAAGATCCTGCAAAAGC 317

Qy        246 TGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGC 305
              |   ||   ||||    |  | | |||| |   |  ||| |  ||   |  |   |||
Db        318 TAGTCAAGGAGGATTAGAAGATCCTGCAAAAGCTAGTCAAGGTGGATTAGAAGATCCTGC 377

Qy        306 CAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 365
              |   |   |||||||  |  |   || | || | |  ||  ||    || |  |   |
Db        378 AAAAGCTAGTCAAGGTGGATTAGAAGATCCCGCAAAGGCTAGTCAAGGTGGATTAGAAGA 437

Qy        366 TAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAA 406
              |  |||||  ||  ||  |||     || |  | ||||
Db        438 TCCTGCAAAAGCTAGTCAAGGTGGATTAGAAGATCCCGCAA 478
```

RESULT 7
ADA31934
ID ADA31934 standard; DNA; 4860 BP.
XX
AC ADA31934;
XX
DT 20-NOV-2003 (first entry)
XX
DE DNA encoding Acinetobacter baumannii protein #3221.
XX
KW ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KW vaccine; plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR P-PSDB; ADA36060.
XX
PT New Acinetobacter baumanii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 3221; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents DNA encoding an A. baumannii
CC protein.
XX
SQ Sequence 4860 BP; 1639 A; 752 C; 998 G; 1471 T; 0 U; 0 Other;

Query Match 8.4%; Score 38.2; DB 8; Length 4860;
Best Local Similarity 45.9%; Pred. No. 0.51;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy        106 AATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAAT 165
              |||   |  |||| |  | |  || | | |||||    | |  |||||| |   || |
Db       3445 AATACTATTGGTAAATTATTAGGTACTGATGCAGCAGATAATCTTCAAGCAGATGCTGAG 3504

Qy        166 AATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGT 225
              | |    ||   |   |  |||||||    |     |    | || ||  ||  | | |
Db       3505 ATTTCAACTATTTATGGTTTGGGAGGCAATGATACGATTCAAGGTGGTGTACAGAATGAT 3564

Qy        226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
              |    | | |    |  ||| | |||  || || |||    |  | |  | | ||  |
Db       3565 TATCTGTATGGTGATGATGGAGACGATACATTAGTTTCTAATACAGGCTCTGATTATTTG 3624

Qy        286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
                ||  ||||    ||| ||   |    | || | | ||   |||     | | |
Db       3625 TATGGTGGTGCAGGTAATGACACATTGATTTATGGCGGTAATTCCAATGTGTATACGGCT 3684

Qy        346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATG 388
              | ||        |||| ||| |  |  |||||| | |  | |
Db       3685 TTACAAGGTCAAGCTGGGAATGACACTTATATTATAGATAAAG 3727
```

RESULT 8
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX

```
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

  Query Match              8.3%;  Score 37.8;  DB 7;  Length 2000;
  Best Local Similarity   13.1%;  Pred. No. 0.47;
  Matches   36;  Conservative  121;  Mismatches  118;  Indels    0;  Gaps    0;

Qy         21 TATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATGATTTAGC 80
              :| |:::: :|  :: |  : ::: :|   : :       : ::  | ::  : :   :
Db        950 WAAGRWRWRMAWCWYCCMWKWWKMTSCMWWKYWRTWSCWYTMWWGAMRYAYYAMRRRRWT 891

Qy         81 TAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGC 140
              : : :  :|:| : :|      :| : :||  | :::  :| : : :|:|||:   ::
Db        890 YKWSWRRMYWTMTKWAWTWMTCMCMAKWYMATGWATWMWWRYTMYTYCYAMTCAKCKYKM 831

Qy        141 CATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACT 200
               :| ::|   ||:: |: ::  :| ::: ::::: | :::  :|:: |: :    : : :
Db        830 AMTKWWTTWACAWRATSWRWRAMAGMRWKRYKMKRAYWWRWWRCWKAGWARWMKSRYRWK 771

Qy        201 TTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGA 260
              :::   : ::: : :  :::::: :   : : :   : :|::       |    :|::|::
Db        770 WKKYATRYYWKMWAMTWWWSWRRWKSYRMWSGMGRMRWSAWRYCSRMKCAKTKYASSARW 711

Qy        261 TTATAACCTTGCATATATTGATCAGGCGGGCAGTG 295
              |::  ::: :: :: :  .:|:|:: :: ::|: |:
Db        710 TKRAKRSYRYRRRWYWKRKGWTYRYRYWRSCRMTR 676


RESULT 9
AAC64623
ID   AAC64623 standard; DNA; 456 BP.
XX
AC   AAC64623;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36347.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

  Query Match              8.2%;  Score 37.4;  DB 3;  Length 456;
  Best Local Similarity   51.5%;  Pred. No. 0.35;
  Matches   86;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              |||   |||   ||   ||  ||||  |||  | |    || |||||  ||  || ||||
Db        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
                | |  ||||| ||||   | ||||| || |  ||    ||||      || ||
Db        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327

Qy        397 AAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGGCTATTCGCGT 443
              ||  | |  |||   ||       | |    ||  ||| ||  || ||
Db        328 AACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGT 374


RESULT 10
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
```

```
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             8.0%;  Score 36.6;  DB 3;  Length 456;
  Best Local Similarity   54.0%;  Pred. No. 0.61;
  Matches   75;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        275 ATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATA 334
              | || |  || || ||   ||  ||||  |||  | |    || |||||  ||  || ||
Db        206 AAATGGCACATGCAGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTA 265

Qy        335 CTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTC 394
              ||       | |  ||||| ||||   | ||||| || |  ||    ||||      ||
Db        266 CTATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTA 325

Qy        395 AAAAAACGGCAATTGTAGT 413
              ||||  | |  |||   ||
Db        326 AAAACTCCGATATTACTGT 344


RESULT 11
AAX34648/c
ID   AAX34648 standard; DNA; 4890 BP.
XX
AC   AAX34648;
XX
DT   17-OCT-2003  (revised)
DT   05-JUL-1999  (first entry)
XX
DE   Promoter containing sequence of SBE I.
XX
KW   Starch biosynthetic pathway; cereal plant; enzyme; SBE; SSS; DBE; GBSS;
KW   starch branching enzyme; starch soluble synthase; debranching enzyme;
KW   endosperm; wheat; barley; granule-bound synthase; glutenin; starch;
KW   grain softness protein I; bacterial isoamylase; glycogen synthase;
KW   wSBE I-D4 gene; promoter; ss.
XX
OS   Aegilops tauschii.
XX
PN   WO9914314-A1.
XX
PD   25-MAR-1999.
XX
PF   11-SEP-1998;   98WO-AU000743.
XX
PR   12-SEP-1997;   97AU-00009108.
PR   20-MAR-1998;   98AU-00002509.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (AUSU ) UNIV AUSTRALIAN NAT.
PA   (GOOD-) GOODMAN FIELDER LTD.
PA   (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.
XX
PI   Li Z,  Morell M,  Rahman S;
XX
DR   WPI; 1999-229525/19.
XX
PT   New isolated cereal plant enzyme genes used for, e.g. expression of
PT   antisense sequences of granule bound synthase.
XX
PS   Claim 17; Page 63-65; 171pp; English.
XX
CC   The invention relates to a novel enzyme of starch biosynthetic pathway in
CC   a cereal plant, where the enzyme is selected from starch branching enzyme
CC   (SBE) I , SBE II, starch soluble synthase (SSS) I, and debranching enzyme
CC   (DBE), with the proviso that the enzyme is not SSS I of rice, or SBE I of
CC   rice or maize. The methods and products can be used for targeting
CC   expression specifically to the endosperm of the seeds of cereal plants
CC   such as wheat or barley. They can be used for the expression of e.g.
CC   antisense sequences of granule-bound synthase (GBSS), SBE II, low mol.
CC   wt. glutenin, grain softness protein I, bacterial isoamylase, bacterial
CC   glycogen synthase, and wheat high mol. wt. glutenin Bx17. They can be
CC   used for modifying the characteristics of starch produced by a plant. The
CC   present sequence represents the wheat SBE I promoter containing sequence.
CC   (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 4890 BP; 1403 A; 1067 C; 1035 G; 1385 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 35.2;  DB 2;  Length 4890;
  Best Local Similarity   55.8%;  Pred. No. 4.3;
  Matches   87;  Conservative    0;  Mismatches   68;  Indels    1;  Gaps    1;

Qy        258 AGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCA 317
              |||  | ||| || |||| | ||| ||  |   |||||||||  |||||| |  || |||
Db       1097 AGAGAAGAACTTTCCATAAAATGAACAAACAAACAGTGCCAAGTATGCCACTTGTTGGCA 1038

Qy        318 AGGTGCTTATGGTAATACTGCGA-TGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATA 376
              |       ||   || | ||  | | |       |  ||||||||  |  |  ||||   |
Db       1037 ACTCATACATTTGAAAATTGTTAGTTAAATGAAACTAAGGTTCTATTCTTTTAGCATGCA 978

Qy        377 TTACACAGTATGGTACTCAAAAAACGGCAATTGTAG 412
              || ||  | ||        | |  ||  ||||||||
Db        977 TTGCATTGCATTACTAGTTAGACCCGTGAATTGTAG 942


RESULT 12
AAQ62647
ID   AAQ62647 standard; DNA; 646 BP.
XX
AC   AAQ62647;
XX
DT   25-MAR-2003  (revised)
DT   18-OCT-1994  (first entry)
XX
DE   Fibronectin binding protein fnbA gene.
XX
KW   Fibronectin binding protein; FBP; curlin; curli pili; pilus;
KW   immunization; vaccine; fnbA gene; ss.
XX
OS   Escherichia coli.
XX
PN   WO9410330-A1.
XX
PD   11-MAY-1994.
XX
PF   03-NOV-1993;   93WO-US010547.
XX
```

```
PR   03-NOV-1992;   92US-00970846.
XX
PA   (REAT/) REA T S.
XX
PI   Normark S,  Olsen A;
XX
DR   WPI; 1994-167484/20.
DR   P-PSDB; AAR52664.
XX
PT   New fibronectin binding protein curlin from E. coli - used for the
PT   treatment of infections caused by E. coli and for immunisation against E.
PT   coli infection.
XX
PS   Disclosure; Fig 2; 33pp; English.
XX
CC   The fnbA gene of E. coli cattle fecal isolate AO12 includes DNA encoding
CC   a new fibronectin binding protein (FBP), curlin, which is produced as
CC   curli pili when expressed in E. coli (pFnb20) DSM 4585. The fnbA gene
CC   sequence from pFnb20 is given in AAQ62647, and includes the sequence
CC   encoding the 17 kDa curlin subunit (amino acids 21-142 of AAR52664).
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 646 BP; 168 A; 146 C; 161 G; 171 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 35;  DB 2;  Length 646;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        416 AGAGACAGTCGCAAATGGCTATTCGCGTGACACAA 450
              |||||||||||||||||||||||||||||||||||
Db          1 AGAGACAGTCGCAAATGGCTATTCGCGTGACACAA 35


RESULT 13
AAQ73066
ID   AAQ73066 standard; DNA; 361 BP.
XX
AC   AAQ73066;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .359
FT                   /*tag=  a
FT                   /note= "AgfA"
FT   misc_feature    37. .60
FT                   /*tag=  d
FT                   /note= "TAF5 primer (pair with TAF6)"
FT   misc_feature    52. .69
FT                   /*tag=  b
FT                   /note= "TAF3 primer (pair with TAF4)"
FT   misc_feature    complement(103. .129)
FT                   /*tag=  e
FT                   /note= "TAF6 primer (pair with TAF5)"
FT   misc_feature    complement(292. .402)
FT                   /*tag=  c
FT                   /note= "TAF4 primer (pair with TAF3)"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR62761.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7A; 95pp; English.
XX
CC   The DNA encodes the Salmonella enteritis27655-3b TnphoA mutant strain
CC   agfA gene cloned into pUC19. The DNA and isolated proteins are used in
CC   genetic immunization and vaccine compositions, respectively, to elicit an
CC   immune response to Salmonella in animals (e.g. food producing animals)
CC   and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             7.6%;  Score 34.6;  DB 2;  Length 361;
  Best Local Similarity   53.3%;  Pred. No. 2.3;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              |||   |||   ||   ||  ||||  |||  | |    || |||||  ||  || ||||
Db        145 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 204

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
                | |  ||||| ||||   | |||||| || |  ||    ||||      || ||
Db        205 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 264

Qy        397 AAAACGGCAATTGTAGT 413
              ||  | |  |||  ||
Db        265 AACTCCGATATTACTGT 281


RESULT 14
AAT74141
ID   AAT74141 standard; DNA; 361 BP.
XX
AC   AAT74141;
XX
DT   25-MAR-2003  (revised)
DT   30-SEP-1997  (first entry)
XX
DE   Salmonella enteritidis 27655-3b TnphoA mutant agfA gene fragment.
XX
KW   Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX
OS   Salmonella enteritidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .360
FT                   /*tag=  a
FT                   /label= agfA_gene_fragment
FT   primer_bind     16. .60
FT                   /*tag=  b
FT                   /label= Primer_TAF5
FT   primer_bind     52. .69
FT                   /*tag=  c
FT                   /label= Primer_TAF3
FT   primer_bind     complement(103. .128)
FT                   /*tag=  d
FT                   /label= Primer_TAF6
FT   primer_bind     complement(294. .312)
FT                   /*tag=  e
FT                   /label= Primer_TAF4
XX
PN   US5635617-A.
```

```
XX
PD   03-JUN-1997.
XX
PF   26-APR-1994;   94US-00233788.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI   Collinson SK,  Kay WW,  Doran JL;
XX
DR   WPI; 1997-309886/28.
DR   P-PSDB; AAW23569.
XX
PT   Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
PT   enteropathogenic bacteria of the Enterobacteria family.
XX
PS   Claim 1; Col 107-110; 85pp; English.
XX
CC   The present sequence represents an isolated agfA gene fragment derived
CC   from Salmonella enteritidis 27655-3b TnphoA mutant strain. The nucleic
CC   acid can be used to provide diagnostic assays for Salmonella and/or
CC   enteropathogenic bacteria of the family Enterobacteria. It can also be
CC   used to provide proteins and antibodies which can be used for assays. The
CC   nucleic acid sequence can be used to provide probes or primers which can
CC   specifically hybridise to nucleic acid molecules from greater than 99% of
CC   Salmonella strains that are pathogenic to warm-blooded animals relative
CC   to nucleic acid molecules from virtually all other microbial organisms.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

  Query Match             7.6%;  Score 34.6;  DB 2;  Length 361;
  Best Local Similarity   53.3%;  Pred. No. 2.3;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              |||   |||   ||   ||  ||||  |||  | |    || |||||  ||  || ||||
Db        145 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 204

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
              | |  ||||| ||||   | ||||| || |  ||    ||||       || ||
Db        205 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 264

Qy        397 AAAACGGCAATTGTAGT 413
              ||  | |  |||   ||
Db        265 AACTCCGATATTACTGT 281


RESULT 15
AAQ87467
ID   AAQ87467 standard; DNA; 456 BP.
XX
AC   AAQ87467;
XX
DT   25-MAR-2003  (revised)
DT   26-JUN-1995  (first entry)
XX
DE   AgfA sequence.
XX
KW   Salmonella; AgfA; vaccine; genetic immunization; ds.
XX
OS   Salmonella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .454
FT                   /*tag= a
FT                   /note= "AgfA"
XX
PN   WO9425598-A2.
XX
PD   10-NOV-1994.
XX
PF   26-APR-1994;   94WO-IB000207.
XX
PR   26-APR-1993;   93US-00054452.
XX
PA   (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA   (KING/) KING J.
XX
PI   Kay WW,  Collinson SK,  Clouthier SC,  Doran JL;
XX
DR   WPI; 1994-358275/44.
DR   P-PSDB; AAR74625.
XX
PT   Eliciting an immune response to Salmonella - using attenuated Salmonella
PT   strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
PS   Disclosure; Fig 7B; 95pp; English.
XX
CC   The DNA encodes the Salmonella AgfA protein. The DNA and isolated
CC   proteins are used in genetic immunization and vaccine compositions,
CC   respectively, to elicit an immune response to Salmonella in animals (e.g.
CC   food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

  Query Match             7.6%;  Score 34.6;  DB 2;  Length 456;
  Best Local Similarity   53.3%;  Pred. No. 2.5;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              |||   |||   ||   ||  ||||  |||  | |    || |||||  ||  || ||||
Db        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
              | |  ||||| ||||   | ||||| || |  ||    ||||       || ||
Db        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327

Qy        397 AAAACGGCAATTGTAGT 413
              ||  | |  |||   ||
Db        328 AACTCCGATATTACTGT 344


Search completed: March 18, 2004, 09:11:52
Job time : 291.405 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 52.4971 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

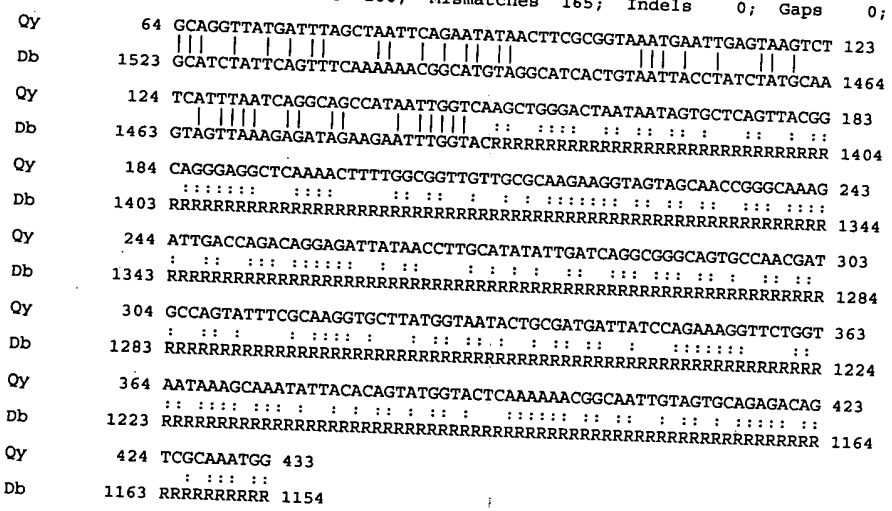
	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	39.6	8.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	38.2	8.4	4860	4	US-09-328-352-3221	Sequence 3221, Ap
	3	34.6	7.6	361	1	US-08-233-788A-56	Sequence 56, Appl
	4	34.6	7.6	456	1	US-08-233-788A-58	Sequence 58, Appl
	5	31.8	7.0	580073	4	US-08-545-528D-1	Sequence 1, Appli
	6	31.8	7.0	640681	4	US-09-790-988-1	Sequence 1, Appli
	7	31.4	6.9	525	4	US-09-107-532A-2325	Sequence 2325, Ap
c	8	31.4	6.9	2785	4	US-08-956-171E-133	Sequence 133, App
	9	31.4	6.9	53332	4	US-09-801-861-3	Sequence 3, Appli
c	10	31.4	6.9	62909	4	US-09-596-002-32	Sequence 32, Appl
	11	31.2	6.8	606	4	US-09-540-236-1518	Sequence 1518, Ap
	12	31.2	6.8	1404	4	US-09-134-001C-398	Sequence 398, App
	13	31	6.8	2139	4	US-09-059-584-50	Sequence 50, Appl
c	14	31	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	15	30.6	6.7	832	4	US-09-621-976-2813	Sequence 2813, Ap
	16	30.6	6.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	17	30.4	6.7	831	4	US-09-328-352-3901	Sequence 3901, Ap
	18	30.4	6.7	1023	4	US-09-540-236-1556	Sequence 1556, Ap
c	19	30.4	6.7	1227	4	US-09-107-532A-2544	Sequence 2544, Ap
c	20	30.4	6.7	1622	4	US-08-956-171E-732	Sequence 732, App
	21	30.4	6.7	1776	4	US-09-134-000C-3324	Sequence 3324, Ap
	22	30.4	6.7	1879	4	US-09-023-655-661	Sequence 661, App
	23	30.4	6.7	2001	3	US-08-714-918-92	Sequence 92, Appl
	24	30.4	6.7	2001	3	US-09-265-315-92	Sequence 92, Appl
	25	30.4	6.7	2001	3	US-09-265-315-92	Sequence 92, Appl
	26	30.4	6.7	2001	3	US-09-266-417-92	Sequence 92, Appl
	27	30.4	6.7	2001	4	US-09-528-709-92	Sequence 92, Appl
	28	30.4	6.7	2001	4	US-09-527-745-92	Sequence 92, Appl
	29	30.4	6.7	99629	4	US-09-596-002-37	Sequence 37, Appl
c	30	30.2	6.6	417	4	US-09-621-976-10111	Sequence 10111, A
	31	30.2	6.6	2364	3	US-09-172-045-1	Sequence 1, Appli
	32	30.2	6.6	2364	4	US-09-342-325C-1	Sequence 1, Appli
	33	30.2	6.6	4215	4	US-09-620-312D-295	Sequence 295, App
	34	30.2	6.6	13508	4	US-08-956-171E-120	Sequence 120, App
	35	30.2	6.6	152331	3	US-09-128-155-16	Sequence 16, Appl
	36	30.2	6.6	176373	3	US-09-128-155-17	Sequence 17, Appl
	37	30	6.6	2136	2	US-08-949-941B-1	Sequence 1, Appli
c	38	29.8	6.5	7091	3	US-08-975-762-46	Sequence 46, Appl
c	39	29.8	6.5	7091	3	US-09-295-028-46	Sequence 46, Appl
c	40	29.8	6.5	7091	4	US-09-106-582-46	Sequence 46, Appl
c	41	29.8	6.5	7091	4	US-09-159-469-46	Sequence 46, Appl
c	42	29.8	6.5	7091	4	US-09-693-542-46	Sequence 46, Appl
c	43	29.8	6.5	13158	2	US-08-687-080-105	Sequence 105, App
c	44	29.4	6.4	703	4	US-09-016-434-120	Sequence 120, App
c	45	29.2	6.4	369	4	US-09-328-352-3808	Sequence 3808, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 8.7%; Score 39.6; DB 1; Length 7218;

Best Local Similarity 10.5%; Pred. No. 0.0088;
Matches 39; Conservative 166; Mismatches 165; Indels 0; Gaps 0;



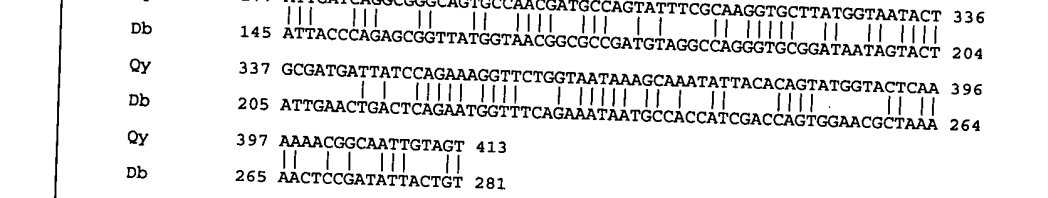
RESULT 2
US-09-328-352-3221
; Sequence 3221, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3221
; LENGTH: 4860
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3221

Query Match 8.4%; Score 38.2; DB 4; Length 4860;
Best Local Similarity 45.9%; Pred. No. 0.021;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;



RESULT 3
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56

Query Match 7.6%; Score 34.6; DB 1; Length 361;
Best Local Similarity 53.3%; Pred. No. 0.087;
Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;



RESULT 4
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.

```
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  456 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..456
US-08-233-788A-58

  Query Match             7.6%;  Score 34.6;  DB 1;  Length 456;
  Best Local Similarity   53.3%;  Pred. No. 0.098;
  Matches   73;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
              ||| |||  ||  ||  |||| ||| | |    || ||||  ||  || ||||
Db        208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Qy        337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
              | |  ||||| ||||  | ||||| || |  ||    ||||       || ||
Db        268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327

Qy        397 AAAACGGCAATTGTAGT 413
              ||  | |  |||  ||
Db        328 AACTCCGATATTACTGT 344


RESULT 5
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
;  APPLICANT: Fraser et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB193P1
;  CURRENT APPLICATION NUMBER: US/08/545,528D
;  CURRENT FILING DATE:  1995-10-19
;  PRIOR APPLICATION NUMBER: US 08/488,018
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/473,545
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 580073
;   TYPE: DNA
;   ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

  Query Match             7.0%;  Score 31.8;  DB 4;  Length 580073;
  Best Local Similarity   43.1%;  Pred. No. 30;
  Matches  153;  Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

Qy         26 TGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATGATTTAGCTAATT 85
              | ||||     |||     | |  | | |||    | ||  | |  | |||||  ||
Db     181203 TTTTAAATGGTTTGGCACTAACAGTTAATTCAGTTTTAACAAATGAACAGTTAGCAGATG 181262

Qy         86 CAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAA 145
              ||     |||||| | | |  |       ||   |  ||  | |||  || | |   ||
Db     181263 CATGTGTTAACTTTGGGTTTGACTTTAAGATGGAAACTGAAGTTACCCATGAAAATATTG 181322

Qy        146 TTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGG 205
              | | | |       |  | |  ||||| | | |||  |||   | | ||   || | |
Db     181323 TAGCTAACATCCAGTTTGAAGATAGTGATGATTTATTGCAACCAAGACCACCTATTGTTA 181382

Qy        206 CGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATA 265
              | | | |  || | |   | | ||     |  ||| |  || |  ||
Db     181383 CTATCATGGGTCATGTTGACCATGGTAAAACTTCGCTTTTAGACACAATTAGAAAAACTA 181442

Qy        266 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 325
              | | |  ||  ||    | || | | || |  |    ||  ||  || |
Db     181443 ATGTAACTGCTAAGGAGTTTGGCGGAATTACCCAAAAAATTGGTGCTTATCAGGTGAAAA 181502

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTAC 380
              ||    | | |  | |||||      | | | ||  |    | | |    ||||
Db     181503 ATCACCAAAATAAAACGATTACTTTTATTGATACTCCTGGGCATGAAGCATTTAC 181557


RESULT 6
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             7.0%;  Score 31.8;  DB 4;  Length 640681;
  Best Local Similarity   46.9%;  Pred. No. 32;
  Matches   99;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          2 TGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCG 61
              | |  ||  ||  || |   || |  ||||||  |       ||   ||  |  |    |
Db      79853 TTAGCAAAGAAACGTCAGAGATAAATTTAACACAAAAAAAAGCATATGAAACAGGTTTCC 79912

Qy         62 CAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGT 121
              | ||||   |    ||  ||    ||  | |  ||   |  |||| |||| | | | |
Db      79913 AAACAGGACTTCTAAAAGGCAAAGAGGAGAACACTTTGTTGAAAAATAAATTAAATGATT 79972
```

```
Qy        122 CTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTAC 181
              |  || |||   | |          ||| | ||  ||  | | |     || |||| ||||
Db      79973 TATTTTAAATTTTGAAAATTCTCTTGCTAAATTTGAAAATGAACTATGTTCTCAATTAC 80032

Qy        182 GGCAGGGAGGCTCAAAACTTTTGGCGGTTGT 212
              |   ||  | |||| |||    |   |||
Db      80033 TAAAAACAGTTTTAAAAATTTCATCTTATGT 80063


RESULT 7
US-09-107-532A-2325
; Sequence 2325, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 2325:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 525 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...525
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2325:
US-09-107-532A-2325

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 525;
  Best Local Similarity   54.9%;  Pred. No. 1.2;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        270 TGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGG 329
              || ||    |||| ||||| |     | ||||| ||||||    |||   |  ||
Db        213 TGAATTATCTGATAAGGCGAAAGAAGAAGCAGTTGCCATTATTTCTTCAGGGATTAGTGA 272

Qy        330 TAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACAC 382
              |||  ||  |||| |    | |    ||.|||  ||||||  ||  |||  ||
Db        273 CTATATTGATATGAGTGGTGAAATCAATTATGGCGATAAAGACAAAATTGTAC 325


RESULT 8
US-08-956-171E-133/c
; Sequence 133, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/956,171E
;             FILING DATE: 20-Oct-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 133:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2785 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 133:
US-08-956-171E-133

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 2785;
  Best Local Similarity   49.7%;  Pred. No. 2.8;
  Matches   80;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        125 CATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGC 184
              || ||||||| ||||  || |||  | || | |   ||  |  ||||| |||| |||
Db       2531 CAATTAATCAAGCAGATATCATTTCTGAACCGGTTGCTCTTGGTAGTGTTCAGGTACCAA 2472

Qy        185 AGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGA 244
              |  ||  ||    |    || |   | ||  ||   ||  ||   |   |       | ||| |
Db       2471 ATGATGGTCAACCTATTATTCTGCTTAATGATAAACAAACGATAGGTGGTTATACCAAAA 2412

Qy        245 TTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
              |||       | |  | | | | ||  || | | | |||
Db       2411 TTGCAACTGTATGCAAATTTGATCTGCCAAAGTTAGCACAG 2371


RESULT 9
US-09-801-861-3
; Sequence 3, Application US/09801861
; Patent No. 6492154
```

```
; GENERAL INFORMATION:
;  APPLICANT:  YAN, Chunhua et al.
;  TITLE OF INVENTION:  ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE:  CL001098
;  CURRENT APPLICATION NUMBER: US/09/801,861
;  CURRENT FILING DATE:  2001-03-09
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 53332
;   TYPE: DNA
;   ORGANISM: Human
US-09-801-861-3

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 53332;
  Best Local Similarity   61.7%;  Pred. No. 12;
  Matches   50;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy        336 TGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCA 395
              ||  |||| ||| ||  |||  | ||   |||||||  ||| ||| |||  | |  |||
Db       8632 TGTTATGAGTATTCAACAAGAATATGTATATAAAGCACTTATCACAGAGTCAGTTTTTCA 8691

Qy        396 AAAAACGGCAATTGTAGTGCA 416
              |    |  ||| ||| |  ||
Db       8692 GAGTTCAACAAATGTTGACCA 8712


RESULT 10
US-09-596-002-32/c
; Sequence 32, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:      Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:       PERL Program
; SEQ ID NO 32
;   LENGTH: 62909
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte template ID No. 6632636 32
;  PUBLICATION INFORMATION:
US-09-596-002-32

  Query Match             6.9%;  Score 31.4;  DB 4;  Length 62909;
  Best Local Similarity   49.7%;  Pred. No. 14;
  Matches   80;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        221 AAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTG 280
              ||   | || |||||       ||  ||     || | ||   | |    ||||||   | |
Db       8029 AATTGATTAACAACCATATGGAGTATGTTATGAAAAGATTATCTTTAATTGCATTATTGG 7970

Qy        281 ATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGA 340
              ||| ||   ||| ||||  || | |  |  ||  | | |||||    |  | ||||
Db       7969 CGGTGGCAGGTTTTGCTAACGCAGCGACTTATGAGCTTGATCCTTATCACACCAATGCGC 7910

Qy        341 TGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACA 381
              ||| |    |  || ||| |   | | ||| |  ||
Db       7909 GTTTTAGCATTGACCATTTTGGCACCTCAACCAATGTGGCA 7869


RESULT 11
US-09-540-236-1518
; Sequence 1518, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2005-001
;  CURRENT APPLICATION NUMBER: US/09/540,236
;  CURRENT FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1518
;   LENGTH: 606
;   TYPE: DNA
;   ORGANISM: M.catarrhalis
US-09-540-236-1518

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 606;
  Best Local Similarity   50.0%;  Pred. No. 1.5;
  Matches   78;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
              | || |||||       ||  ||    || | ||   | |    ||||||   | |    |
Db          4 ATTAACAACCATATGGAGTATGTTATGAAAAGATTATCTTTAATTGCATTATTGGCGGTG 63

Qy        286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
              || ||   ||| ||||  || | |  |  ||  | | |||||    |  | ||||   ||
Db         64 GCAGGTTTTGCTAACGCAGCGACTTATGAGCTTGATCCTTATCACACCAATGCGCGTTTT 123

Qy        346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACA 381
              | |    |   || ||| |    | | ||| |  ||
Db        124 AGCATTGACCATTTTGGCACCTCAACCAATGTGGCA 159


RESULT 12
US-09-134-001C-398
; Sequence 398, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 398
;   LENGTH: 1404
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-398

  Query Match             6.8%;  Score 31.2;  DB 4;  Length 1404;
  Best Local Similarity   52.3%;  Pred. No. 2.3;
  Matches   69;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        275 ATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATA 334
              ||  |||| ||   |   ||   ||  ||  | || |||   | | ||| | | || |
Db       1079 ATTCTGATAAGAACGCAGATGATGACTCTGATAATAGTTCCTCAAATCCTAACGCTACTT 1138

Qy        335 CTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTC 394
              |  | |  | || | | || |   |   ||||||  |||||  ||||  |  |   |  |
Db       1139 CAACTAATAATAACGATAATGTAGCCAATAATAACTCAAATTATACAAACCAAAATCAAC 1198

Qy        395 AAAAAACGGCAA 406
              || | |  ||||
Db       1199 AAGATAATGCAA 1210
```

```
RESULT 13
US-09-059-584-50
; Sequence 50, Application US/09059584
; Patent No. 6440701
;  GENERAL INFORMATION:
;    APPLICANT:  Myers, Lisa E
;    APPLICANT:  Schryvers, Anthony B
;    APPLICANT:  Harkness, Robin E
;    APPLICANT:  Loosmore, Sheena M.
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Yang, Yan-Ping
;    APPLICANT:  Klein, Michel H
;    TITLE OF INVENTION:  Transferrin Receptor Genes of Moraxella
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/059,584
;      FILING DATE:  14-APR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/778,570
;      FILING DATE:  03-JAN-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24973
;      REFERENCE/DOCKET NUMBER:  1038-794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2139 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-059-584-50

  Query Match             6.8%;  Score 31;  DB 4;  Length 2139;
  Best Local Similarity   59.8%;  Pred. No. 3.3;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy        306 CAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 365
              ||||  |    ||  ||||| || || |||||| |  |||      || |||| | ||
Db       1845 CAGTGATGGTAAATCTGCTTTTGATATTACTGCTACAATTGATGGCAATGGTTTTAGTGG 1904

Qy        366 TAAAGCAAATATTACACAGTATGGTAC 392
              |||||| ||||   || |   ||  ||
Db       1905 TAAAGCCAATACACCAGATATTGAAAC 1931


RESULT 14
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;  LENGTH: 1664976
;  TYPE: DNA
;  ORGANISM: Methanococcus jannaschii
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (28222)..(28222)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (28257)..(28258)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (84773)..(84773)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (84808)..(84808)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (84812)..(84812)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (98120)..(98120)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (98159)..(98159)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (98239)..(98239)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (98266)..(98266)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (98343)..(98343)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (103998)..(103998)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (148948)..(148948)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (163385)..(163385)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (191989)..(191989)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (191995)..(191995)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (231980)..(231980)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (234187)..(234187)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (234220)..(234220)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (234814)..(234814)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (309398)..(309398)
;  OTHER INFORMATION: n equals a, t, c, or g
;  NAME/KEY: misc_feature
;  LOCATION: (309418)..(309418)
;  OTHER INFORMATION: n equals a, t, c, or g
```

; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 6.8%; Score 31; DB 4; Length 1664976;
Best Local Similarity 49.1%; Pred. No. 83;
Matches 82; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy      239 CAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCA 298
            || | | | |  ||   |  ||   || | | ||   |  ||| || || || ||  |
Db   122859 CATATAATAAGGAGGATGATGAAACTACCGTGGCTACTTATGACCATGCAGGGAGGACTG 122800

Qy      299 ACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTT 358
              |||  || | | | |   |  | ||| | |  |||      |  ||   || | |
Db   122799 GAGATTACAATTATGCAAATAATATTGATGCTTACGCTGGTTCAAAAATAGCGATAAAAT 122740

Qy      359 CTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCA 405
             |  | |  ||||  |  | |||||||||  |  |  ||| |||
Db   122739 ATACAAGTGGAGCATTAACAAGACAGTATGGGGTTTTAGGAACTGCA 122693
```

RESULT 15
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match 6.7%; Score 30.6; DB 4; Length 832;
Best Local Similarity 16.5%; Pred. No. 2.8;
Matches 69; Conservative 151; Mismatches 195; Indels 2; Gaps 1;

```
Qy       10 AAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGT 69
            ::::::: :::::: :| :::::|::  ::::  |:  |::: ::|:: :      ::
Db       40 WRRKKKKAWWKYWKTWTWWYWRYAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYM 99

Qy       70 TATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTT 129
             :: :|:|   :  :  :  :::: :  :  :  :  |    : :|:  ::  :  |: :::|
Db      100 AMWTRTWTGYAYYRSMMYWWRYRCWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKT 159
```

```
Qy        130 AATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGA 189
Db        160 YYWAATRYWWMMCWTKRWRASWWYCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGA 219

Qy        190 GGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGAC 249
Db        220 MSWKYMWRMWRWRGWATGAGMKAWRASCMMRRKYAGKSKTSYKSMWMCWTRSWKYCYTKA 279

Qy        250 CAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGT 309
Db        280 RWTGYYCYRKGGMWGKRGRWYASKKYMWKRWWWCWARMYRYSTGTRASMWWRRWYYTMMM 339

Qy        310 ATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCA--GAAAGGTTCTGGTAATA 367
Db        340 KWWKYAWARAAWRWWAMWWAWRRACAAAATATAATTATTTATGGTACAATTCTTGTACTT 399

Qy        368 AAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGT 424
Db        400 TAGCAAATCTGGAGTTAGTTCATAGTCAAAGTCAGTTAATATTTCTTAGAGGAAAGT 456

Search completed: March 18, 2004, 20:20:09
Job time : 58.4971 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 250.734 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	36	7.9	413	14	US-10-184-644-156	Sequence 156, App
	2	36	7.9	413	14	US-10-184-634-156	Sequence 156, App
c	3	34.4	7.5	5184	14	US-10-247-671-55	Sequence 55, Appl
	4	34	7.5	12426	12	US-10-221-714A-204	Sequence 204, App
c	5	33.8	7.4	1093	12	US-10-424-599-43422	Sequence 43422, A
	6	33.8	7.4	1149	9	US-09-912-020-139	Sequence 139, App
	7	33.8	7.4	1149	12	US-10-282-122A-6441	Sequence 6441, Ap
	8	33.6	7.4	1299	12	US-10-282-122A-36109	Sequence 36109, A
	9	33.2	7.3	1371	12	US-10-282-122A-17259	Sequence 17259, A
	10	33	7.2	312	12	US-10-424-599-104721	Sequence 104721,
c	11	33	7.2	1194	15	US-10-369-493-32803	Sequence 32803, A
	12	33	7.2	1269	12	US-10-282-122A-35991	Sequence 35991, A
	13	33	7.2	1837	12	US-10-282-122A-35601	Sequence 35601, A
c	14	33	7.2	319630	15	US-10-398-221-7	Sequence 7, Appli
	15	33	7.2	3011208	15	US-10-398-221-2058	Sequence 2058, Ap
c	16	32.8	7.2	746	15	US-10-027-632-14807	Sequence 14807, A
c	17	32.8	7.2	746	15	US-10-027-632-14808	Sequence 14808, A
	18	32.8	7.2	2552	15	US-10-027-632-111939	Sequence 111939,
	19	32.6	7.1	351	9	US-09-983-965-4342	Sequence 4342, Ap
c	20	32.6	7.1	2000	15	US-10-260-238-1758	Sequence 1758, Ap
c	21	32.6	7.1	2250	15	US-10-369-493-46046	Sequence 46046, A
c	22	32.6	7.1	17419	14	US-10-239-676-100	Sequence 100, App
c	23	32.6	7.1	17419	14	US-10-311-455-1268	Sequence 1268, Ap
c	24	32.6	7.1	17419	14	US-10-240-453-112	Sequence 112, App
	25	32.4	7.1	1320	12	US-10-282-122A-10004	Sequence 10004, A
	26	32.4	7.1	2379	12	US-10-282-122A-9449	Sequence 9449, Ap
	27	32.2	7.1	2634	12	US-10-282-122A-35828	Sequence 35828, A
	28	32	7.0	2565	15	US-10-369-493-25168	Sequence 25168, A
c	29	31.8	7.0	409	9	US-09-864-761-2955	Sequence 2955, Ap
c	30	31.8	7.0	438	9	US-09-864-761-19736	Sequence 19736, A
	31	31.8	7.0	708	14	US-10-184-644-584	Sequence 584, App
	32	31.8	7.0	708	14	US-10-184-634-584	Sequence 584, App
	33	31.8	7.0	1860	12	US-10-282-122A-27280	Sequence 27280, A
	34	31.8	7.0	11138	10	US-09-960-870-5	Sequence 5, Appli
	35	31.8	7.0	11138	10	US-09-960-858-5	Sequence 5, Appli
	36	31.8	7.0	580073	14	US-10-205-220-1	Sequence 1, Appli
	37	31.8	7.0	640681	9	US-09-790-988-1	Sequence 1, Appli
	38	31.6	6.9	480	9	US-09-815-242-4053	Sequence 4053, Ap
	39	31.6	6.9	480	12	US-10-282-122A-7419	Sequence 7419, Ap
c	40	31.6	6.9	575	15	US-10-027-632-222766	Sequence 222766,
	41	31.6	6.9	2565	15	US-10-321-802-19	Sequence 19, Appl
c	42	31.4	6.9	867	15	US-10-027-632-167010	Sequence 167010,
c	43	31.4	6.9	867	15	US-10-027-632-167011	Sequence 167011,
c	44	31.4	6.9	867	15	US-10-027-632-167012	Sequence 167012,
c	45	31.4	6.9	1407	12	US-10-424-599-25067	Sequence 25067, A

ALIGNMENTS

RESULT 1
US-10-184-644-156
; Sequence 156, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 156
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-156

Query Match 7.9%; Score 36; DB 14; Length 413;
Best Local Similarity 11.6%; Pred. No. 2;
Matches 28; Conservative 93; Mismatches 121; Indels 0; Gaps 0;

```
Qy      147 TGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGC 206
            ::|:  ::  :: ::   :::  : | : :| ::::: |:||: :   : :     :||
Db      172 HDGKDYVKGSKKLRVGLLKMRNKSGGKKRRGSKRSRREASGGDQREGTREHLQERAKGGR 231

Qy      207 GGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAA 266
```

```
Db        232 RRKKSGRGQRIAEGRPSFQWTRVKNTHIPKGWARGGMGDATLDYDYALLELKRAHKKKYM 291

Qy        267 CCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTA 326

Db        292 ELGISPTIKKMPGGMIHFSGFDNDRADQLVYRFCSVSDESNDLLYQYCDAESGSTGSGVY 351

Qy        327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386

Db        352 LRLKDPDKKNWKRKIIAVYSGHQWVDVHGVQKDYNVAVRITPLKYAQICLWIHGNDANCA 411

Qy        387 TG 388

Db        412 YG 413
```

RESULT 2
US-10-184-634-156
; Sequence 156, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 156
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-156

Query Match 7.9%; Score 36; DB 14; Length 413;
Best Local Similarity 11.6%; Pred. No. 2;
Matches 28; Conservative 93; Mismatches 121; Indels 0; Gaps 0;

```
Qy        147 TGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGC 206

Db        172 HDGKDYVKGSKKLRVGLLKMRNKSGGKKRRGSKRSRREASGGDQREGTREHLQERAKGGR 231

Qy        207 GGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAA 266

Db        232 RRKKSGRGQRIAEGRPSFQWTRVKNTHIPKGWARGGMGDATLDYDYALLELKRAHKKKYM 291

Qy        267 CCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTA 326

Db        292 ELGISPTIKKMPGGMIHFSGFDNDRADQLVYRFCSVSDESNDLLYQYCDAESGSTGSGVY 351

Qy        327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386

Db        352 LRLKDPDKKNWKRKIIAVYSGHQWVDVHGVQKDYNVAVRITPLKYAQICLWIHGNDANCA 411

Qy        387 TG 388

Db        412 YG 413
```

RESULT 3
US-10-247-671-55/c
; Sequence 55, Application US/10247671
; Publication No. US20030194721A1
; GENERAL INFORMATION:
; APPLICANT: Mikita, Thomas
; APPLICANT: Shiffman, Dov
; APPLICANT: Porter, Gordon, J.
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
; FILE REFERENCE: PA-0050 US
; CURRENT APPLICATION NUMBER: US/10/247,671
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/323,784
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PERL Program
; SEQ ID NO 55
; LENGTH: 5184
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030194721A1 5185743CB1
US-10-247-671-55

Query Match 7.5%; Score 34.4; DB 14; Length 5184;
Best Local Similarity 52.9%; Pred. No. 23;
Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy         51 GATTGCAGCCGCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGA 110

Db       2489 GCTTGCAGCTGCTGAATAGTATTCTGAAGCTGACTCACACTATTTGGTGATGGAAAAATA 2430

Qy        111 ATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAG 170

Db       2429 TTTGATGAAATCTGCTGCTGTAAAGTCTGTGCTTGTTCTTGTAGTGGTGACTGCTGCTGT 2370

Qy        171 TGCTCAGTTACGGCAGGGAG 190

Db       2369 TGTTGTGATGCCTCAGTCAG 2350
```

RESULT 4
US-10-221-714A-204
; Sequence 204, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 204
; LENGTH: 12426
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-204

```
  Query Match             7.5%;  Score 34;  DB 12;  Length 12426;
  Best Local Similarity   48.5%;  Pred. No. 50;
  Matches   94;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy         14 TGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATG 73
              ||||  ||||  || |    | | |    |||  | | |||  |   | ||    || |
Db       9253 TGTTTGTTATTGTGGTTTATAGATTTTAAGCGATTTGTATTTAAATGGTAGAGTTTTTTT 9312

Qy         74 ATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATC 133
              ||||    |  || |   | ||| ||      ||||   ||   | ||| ||  |||||||
Db       9313 ATTTTTTTTTTTGAAGTTTTAATTTTTTTTAAAATATTTTTGTTTAGTGTTTGTTTAATC 9372

Qy        134 AGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCT 193
              ||      | |||  | | | | ||| |   |  |    | ||  |        | |   |
Db       9373 GGGGTAGAAATATTTTTAATGATAGGAATTTAATAATATTTAAGGAAGTTTTAGTAAAGT 9432

Qy        194 CAAAACTTTTGGCG 207
              |||  |||| | |
Db       9433 GAAATGTTTTTGGG 9446


RESULT 5
US-10-424-599-43422/c
; Sequence 43422, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 43422
;   LENGTH: 1093
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_139209C.1
US-10-424-599-43422

  Query Match             7.4%;  Score 33.8;  DB 12;  Length 1093;
  Best Local Similarity   58.4%;  Pred. No. 15;
  Matches   59;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        294 TGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAA 353
              ||| |||  | | |      ||  |||||| | | | | || | | ||  |||  | || |
Db        640 TGCAAACAGTTCGATGTCATCAGAAGGTGATAAAGTTTATCCCGAGAGTATTTGCAAGGA 581

Qy        354 AGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTC 394
              || |||| ||  |||  |     |||||    |||||| ||
Db        580 AGATTCTTGTGTTAAGTCTGCAGTTACAGCTAATGGTAGTC 540


RESULT 6
US-09-912-020-139
; Sequence 139, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Trawick, John
;  APPLICANT:  Forsyth, R. Allyn
;  APPLICANT:  Froelich, Jamie M.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001DV1
;  CURRENT APPLICATION NUMBER: US/09/912,020
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 139
;   LENGTH: 1149
;   TYPE: DNA
;   ORGANISM: E. Coli
US-09-912-020-139

  Query Match             7.4%;  Score 33.8;  DB 9;  Length 1149;
  Best Local Similarity   56.9%;  Pred. No. 16;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        259 GATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAA 318
              ||| |||||||    ||||  | |||||  || |   || |||  |  | ||  |||
Db        442 GATAATAACCTGAATTATAGCGTTCAGGTCGGTAACACCCACGGAGGTAATACATCGTCT 501

Qy        319 GGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATA 367
              ||  |   |||| | | | |  | | |    |    || || ||||||||
Db        502 GGCACCAGTGGTTACAGTTCTCTTAATTATCGTGGAGCTTATGGTAATA 550


RESULT 7
US-10-282-122A-6441
; Sequence 6441, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6441
```

```
;   LENGTH: 1149
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-282-122A-6441

  Query Match             7.4%;  Score 33.8;  DB 12;  Length 1149;
  Best Local Similarity   56.9%;  Pred. No. 16;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        259 GATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAA 318
              ||| ||||||||    ||||  | |||||  || |    || |||  |  | ||  |||
Db        442 GATAATAACCTGAATTATAGCGTTCAGGTCGGTAACACCCACGGAGGTAATACATCGTCT 501

Qy        319 GGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATA 367
              ||  |   |||| | | | |  | | |    |    || || ||||||||
Db        502 GGCACCAGTGGTTACAGTTCTCTTAATTATCGTGGAGCTTATGGTAATA 550


RESULT 8
US-10-282-122A-36109
; Sequence 36109, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36109
;   LENGTH: 1299
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
US-10-282-122A-36109

  Query Match             7.4%;  Score 33.6;  DB 12;  Length 1299;
  Best Local Similarity   59.4%;  Pred. No. 19;
  Matches   57;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy        299 ACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTT 358
              | |||   |||| | ||||||  ||||||| | |  | |  | || | |  | || | |
Db        314 ATGATAATAGTACTGCGCAAGCAGCTTATGATCAAGCAGTCAGGAATCTAGATAAGGTTA 373

Qy        359 CTGGTAATAAAGCAAATATTACACAGTATGGTACTC 394
              || | |   |||||  || |   |||||||| |
Db        374 ACCGTCAAATTACAAATTATAAAACTTATGGTACGC 409


RESULT 9
US-10-282-122A-17259
; Sequence 17259, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17259
;   LENGTH: 1371
;   TYPE: DNA
;   ORGANISM: Clostridium difficile
US-10-282-122A-17259

  Query Match             7.3%;  Score 33.2;  DB 12;  Length 1371;
  Best Local Similarity   48.0%;  Pred. No. 26;
  Matches   95;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy         82 AATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCC 141
              |||  |||| || | |  || ||||  ||  || ||| ||||  ||| ||| |   |
Db        466 AATAAAGAACATGAGTGTGCTGTAATAGAGATGGGTATGTCTGGATTCAATGAAATAAAA 525

Qy        142 ATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTT 201
              | | |  |  | | | ||| |.|||     || | |   ||| | | || |
Db        526 TATTTAGTGGATATAGTAAATCCTAAAATTGCAGTCATATCGAATATAGGATTATCACAT 585

Qy        202 TTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGAT 261
              | |     | |   | |||  |||          |||||||| ||    ||  | |
Db        586 GTAGAAAAACTAGGTTCACAAGAGGGTATTTTGAAAGCAAAGATGGAGATTACTTCAAAC 645
```

```
Qy        262 TATAACCTTGCATATATT 279
              | | |     || ||| |
Db        646 TTTGATGAAACAAATACT 663
```

RESULT 10
US-10-424-599-104721
; Sequence 104721, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 104721
; LENGTH: 312
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_65579C.1
US-10-424-599-104721

Query Match 7.2%; Score 33; DB 12; Length 312;
Best Local Similarity 52.6%; Pred. No. 13;
Matches 72; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        267 CCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTA 326
              || ||| |  | |||| |  |   | |||||   | |||  |  |||| |||| ||  |
Db        129 CCATGCTTTGAATGATAAAACTTACCGTGCCCGTGCTGCATGGTTTTCCCAAGCGGCAAA 188

Qy        327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386
                || |    |  ||||  ||| |   ||  ||||||  ||| |    |    || |
Db        189 AAAAAAAAAAAAGGGGATTGGCCATACGAGTCATGGTAACCAAGAGAGACATGATGAGCA 248

Qy        387 TGGTACTCAAAAAACGG 403
               | ||  | | | || |
Db        249 CGATAAACGATATACTG 265
```

RESULT 11
US-10-369-493-32803/c
; Sequence 32803, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 32803
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Chloroflexus aurantiacus
US-10-369-493-32803

Query Match 7.2%; Score 33; DB 15; Length 1194;
Best Local Similarity 48.6%; Pred. No. 28;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```
Qy        167 ATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTA 226
              | | ||  | |||     ||| ||  | | ||  ||  | |  |||     |||  |||||
Db        737 ACACTGTCCGGTTCGACCAGCGAATCACCAACATTCAGCCCTTTGGCAAGCAGAAAGGTG 678

Qy        227 GTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGG 286
              | || |     | |  | ||| |||       ||   || ||        | | ||||| |
Db        677 TTGGCCATACCACCACCAATGAGCAGTGCATCGACCTTACCCAGCAGATTTTCGATCACG 618

Qy        287 CGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTA 346
              | |    || ||  |||    | |   | | |  | ||| |     ||   |  |||||
Db        617 CCGATCTTGTCACTGATTTTCGCACCACCGATGACGGTTACGAATGGACGCTGCGGATTA 558

Qy        347 TCCAG 351
              | |||
Db        557 TTCAG 553
```

RESULT 12
US-10-282-122A-35991
; Sequence 35991, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35991
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Streptococcus mutans
US-10-282-122A-35991

Query Match 7.2%; Score 33; DB 12; Length 1269;
Best Local Similarity 49.2%; Pred. No. 29;
Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```
Qy        236 GGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTG 295
```

```
Db        413 GAGCAAAGATTCAACAGACAGGTGGTGATATTACAGCCACTGCCGATCGTTTAAAGGGTG 472

Qy        296 CCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAG 355

Db        473 CTAACATTTACATGGATTTTCCAAGTGTTGGCGCGACTCAGAATCTTATGATGGCTGCTA 532

Qy        356 GTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAG 412

Db        533 CTTTAGCTGATGGAACAACGATTATCGAAAATGCAGCACGTGAACCAGAAATCGTTG 589
```

RESULT 13
US-10-282-122A-35601
; Sequence 35601, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35601
; LENGTH: 1837
; TYPE: DNA
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35601

Query Match 7.2%; Score 33; DB 12; Length 1837;
Best Local Similarity 47.8%; Pred. No. 35;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy        197 AACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAG 256

Db       1449 AAATAGAGACTGATACTGGTAAACGATTAATTATTAAACTTGAAACAATTAGTGAGCCGG 1508

Qy        257 GAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGC 316

Db       1509 ATGAAAATGGTTATCGAACAATTTATTATGTTATGAATGGTCAAGCACGTCGTATTTCTA 1568

Qy        317 AAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATA 376

Db       1569 TAAAAGATGAAAATATTAAAACAAATACGAACTTGAAACCTAAAGCTGACAAATCAAATC 1628

Qy        377 TTACACAGTATGGTACTCAAA 397

Db       1629 CTAGTCATATTGGTGCTCAAA 1649
```

RESULT 14
US-10-398-221-7/c
; Sequence 7, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 319630
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-7

Query Match 7.2%; Score 33; DB 15; Length 319630;
Best Local Similarity 52.6%; Pred. No. 5.6e+02;
Matches 72; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        249 CCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAG 308

Db     285383 CCCTTCACGAGTTCATAGCCTTTCGTATTTCGATATATAGTTTTGAAATATAAATCACCA 285324

Qy        309 TATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAA 368

Db     285323 TATTTCAAAACTATTGTATGATAATTCGGCTATTAAAGTCAAGTTATTTTGAAATATAGG 285264

Qy        369 AGCAAATATTACACAGT 385

Db     285263 TGCTAATATTTCAAAAT 285247
```

RESULT 15
US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208

```
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2058

  Query Match             7.2%;  Score 33;  DB 15;  Length 3011208;
  Best Local Similarity   52.6%;  Pred. No. 9.5e+02;
  Matches   72;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        249 CCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAG 308
              ||   || ||| | ||| |||| | ||| | |||     |    |    |   ||  |
Db    1832040 CCCTTCACGAGTTCATAGCCTTTCGTATTTCGATATATAGTTTTGAAATATAAATCACCA 1832099

Qy        309 TATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAA 368
              ||||||  ||       |||| |||| | || || |   || ||  |  ||    ||
Db    1832100 TATTTCAAAACTATTGTATGATAATTCGGCTATTAAAGTCAAGTTATTTTGAAATATAGG 1832159

Qy        369 AGCAAATATTACACAGT 385
               || |||||| || | |
Db    1832160 TGCTAATATTTCAAAAT 1832176


Search completed: March 19, 2004, 03:48:55
Job time : 257.734 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 2650.77 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-4	Sequence 4, Appli
2	332.8	73.0	456	23	US-09-543-407-2	Sequence 2, Appli
3	290.2	63.6	561	17	US-09-252-691-171	Sequence 171, App
4	290.2	63.6	561	17	US-09-252-691C-171	Sequence 171, App
5	290.2	63.6	561	51	US-10-417-886-171	Sequence 171, App
6	100	21.9	100	45	US-10-146-492B-80	Sequence 80, Appl
7	100	21.9	100	45	US-10-146-492B-81	Sequence 81, Appl
8	100	21.9	100	45	US-10-146-492B-82	Sequence 82, Appl
c 9	39.6	8.7	7218	8	US-08-466-194-14	Sequence 14, Appl
10	38.2	8.4	391	18	US-09-332-782-15700	Sequence 15700, A
11	38.2	8.4	391	21	US-09-515-694-15700	Sequence 15700, A
12	38.2	8.4	4860	51	US-10-431-652-3221	Sequence 3221, Ap
c 13	37.8	8.3	2000	36	US-09-887-272A-5263	Sequence 5263, Ap
14	37.4	8.2	456	23	US-09-543-407-13	Sequence 13, Appl
c 15	37.4	8.2	68667	39	US-09-949-016-17017	Sequence 17017, A
c 16	37.4	8.2	76668	103	US-60-466-412-82785	Sequence 82785, A
c 17	36.8	8.1	573	43	US-10-021-323-3919	Sequence 3919, Ap
c 18	36.8	8.1	573	82	US-60-255-619-3919	Sequence 3919, Ap
19	36.6	8.0	456	23	US-09-543-407-25	Sequence 25, Appl
c 20	36.4	8.0	396071	103	US-60-465-241-51735	Sequence 51735, A
21	36.2	7.9	42918	103	US-60-466-412-85430	Sequence 85430, A
22	36.2	7.9	2105303	39	US-09-947-916-273	Sequence 273, App
23	36	7.9	413	46	US-10-176-912-156	Sequence 156, App
24	36	7.9	413	46	US-10-179-524-156	Sequence 156, App
25	36	7.9	413	46	US-10-184-634-156	Sequence 156, App
26	36	7.9	413	46	US-10-184-644-156	Sequence 156, App
c 27	36	7.9	395143	103	US-60-466-412-87989	Sequence 87989, A
c 28	36	7.9	395143	106	US-60-499-964-5135	Sequence 5135, Ap
29	35.8	7.9	85690	29	US-09-692-412-72	Sequence 72, Appl
30	35.8	7.9	85690	33	US-09-803-736-1052	Sequence 1052, Ap
31	35.8	7.9	85690	55	US-10-746-294-72	Sequence 72, Appl
32	35.6	7.8	9493	70	US-60-144-351-3079	Sequence 3079, Ap
33	35.6	7.8	12178	26	US-09-620-392-28400	Sequence 28400, A
34	35.6	7.8	12178	30	US-09-702-134-10097	Sequence 10097, A
35	35.6	7.8	12178	33	US-09-815-264-80300	Sequence 80300, A
c 36	35.6	7.8	13036	26	US-09-620-392-47103	Sequence 47103, A
c 37	35.6	7.8	13036	30	US-09-702-134-22694	Sequence 22694, A
c 38	35.6	7.8	13036	33	US-09-815-264-64696	Sequence 64696, A
39	35.6	7.8	32948	26	US-09-620-392-40033	Sequence 40033, A
40	35.6	7.8	32948	30	US-09-702-134-20486	Sequence 20486, A
41	35.6	7.8	32948	33	US-09-815-264-66935	Sequence 66935, A
c 42	35.6	7.8	59719	39	US-09-949-016-15616	Sequence 15616, A
c 43	35.6	7.8	163183	107	US-60-500-337-19909	Sequence 19909, A
44	35.4	7.8	610	44	US-10-098-754-5821	Sequence 5821, Ap
45	35.4	7.8	837105	39	US-09-948-124-76	Sequence 76, Appl

ALIGNMENTS

RESULT 1
US-09-543-407-4
; Sequence 4, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. Coli

US-09-543-407-4

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 9e-125;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              ||||||||||||||||||||||||||||||||||||
Db        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
```

RESULT 2
US-09-543-407-2
; Sequence 2, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Salmonella enteritidis
US-09-543-407-2

Query Match 73.0%; Score 332.8; DB 23; Length 456;
Best Local Similarity 83.1%; Pred. No. 4.7e-88;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db          1 ATGAAAAACAAATTGTTATTTATGATGTTGACAATACTGGGTGCGCCTGGGATTGCAACC 60

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||  ||  |||||||  | |||  |||||||||||| || |||||||||||||| || |||
Db         61 GCGACAAATTATGATCTGGCTCGTTCAGAATATAATTTTGCGGTAAATGAATTAAGCAAG 120

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
```

```
                ||||||||||||||||||  |||||  ||||||||||  ||  ||  ||||||||||      ||
Db        121 TCTTCATTTAATCAGGCGGCCATTATTGGTCAAGTCGGCACGGATAATAGTGCCAGAGTA 180

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
                || ||||  |||  ||||||||  |||  |  |||  ||  |||||||||   |  |   ||  |||||
Db        181 CGCCAGGAAGGATCAAAACTATTGTCCGTTATTTCACAAGAAGGAGGAAATAATCGGGCG 240

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
                ||   |  ||||||  ||||   |||||||| ||||  ||||||||  ||   ||||||  |||||||
Db        241 AAAGTCGACCAGGCAGGGAATTATAACTTTGCGTATATTGAGCAAACGGGCAATGCCAAC 300

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
                ||||||||||  ||||||  |  |||||  |||||||  |||     ||||||||||||||||||
Db        301 GATGCCAGTATATCGCAAAGCGCTTACGGTAATAGTGCAGCTATTATCCAGAAAGGTTCT 360

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
                ||  |||||  ||  ||||||||  |||||  |||||  ||  |||||  |||  ||||||||||||  |
Db        361 GGAAATAAGGCCAATATTACCCAGTACGGTACGCAGAAAACAGCAGTTGTAGTGCAGAAA 420

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
                ||||||||  ||||||||||||||  ||  |||||  |||
Db        421 CAGTCGCATATGGCTATTCGCGTCACCCAACGCTAA 456


RESULT 3
US-09-252-691-171
; Sequence 171, Application US/09252691B
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691B
;  CURRENT FILING DATE:  1999-02-18
;  NUMBER OF SEQ ID NOS:  11324
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691-171

  Query Match             63.6%;  Score 290.2;  DB 17;  Length 561;
  Best Local Similarity   77.4%;  Pred. No. 2.5e-75;
  Matches  352;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
                ||||||||||||  |||  ||||||||||| ||| ||||||||||||||||||  |  |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
                ||||  | ||  |||||||| | |||  |||||||| || |||  |||||||||| |||||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
                 ||||||  ||||||||||||||  |||||||||    ||  |    ||||  | ||  |  ||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
                ||  ||||   ||  ||  |||  |   ||  |  ||  ||  ||   |  |   |||||  ||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
                ||  ||||  |||  ||||     ||||||||||  |||||  ||||||  |  ||||  |  ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
                |||||  ||||||   |||  ||  ||||  |||  ||  ||  ||  ||||||||||||||||||  ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
                |||||  |  ||  |||||  ||  ||||||||||  ||  |||||  |||  |||||||  ||||||
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTGCAGAGA 525

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
                ||||||||||||||||||||||  |  ||||||||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTAA 560


RESULT 4
US-09-252-691C-171
; Sequence 171, Application US/09252691C
; GENERAL INFORMATION:
;  APPLICANT:  Keith G. Weinstock et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.135
;  CURRENT APPLICATION NUMBER: US/09/252,691C
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,145
;  PRIOR FILING DATE:  1998-07-24
;  PRIOR APPLICATION NUMBER:  US 60/074,787
;  PRIOR FILING DATE:  1998-02-18
;  NUMBER OF SEQ ID NOS:  11326
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-09-252-691C-171

  Query Match             63.6%;  Score 290.2;  DB 17;  Length 561;
  Best Local Similarity   77.4%;  Pred. No. 2.5e-75;
  Matches  352;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
                ||||||||||||  |||  ||||||||||| ||| ||||||||||||||||||  |  |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
                ||||  | ||  |||||||| | |||  |||||||| || |||  |||||||||| |||||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
                 ||||||  ||||||||||||||  |||||||||    ||  |    ||||  | ||  |  ||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
                ||  ||||   ||  ||  |||  |   ||  |  ||  ||  ||   |  |   |||||  ||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
                ||  ||||  |||  ||||     ||||||||||  |||||  ||||||  |  ||||  |  ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
                |||||  ||||||   |||  ||  ||||  |||  ||  ||  ||  ||||||||||||||||||  ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
                |||||  |  ||  |||||  ||  ||||||||||  ||  |||||  |||  |||||||  ||||||
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTA 455
                ||||||||||||||||||||||  |  ||||||||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 5
US-10-417-886-171
; Sequence 171, Application US/10417886
; GENERAL INFORMATION:
;  APPLICANT: Keith G. Weinstock et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
;  TITLE OF INVENTION:  CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.135
;  CURRENT APPLICATION NUMBER: US/10/417,886
;  CURRENT FILING DATE:  2003-04-17
```

```
;  PRIOR APPLICATION NUMBER: US/09/252,691C
;  PRIOR FILING DATE: 1999-02-18
;  PRIOR APPLICATION NUMBER: US 60/094,145
;  PRIOR FILING DATE: 1998-07-24
;  PRIOR APPLICATION NUMBER: US 60/074,787
;  PRIOR FILING DATE: 1998-02-18
;  NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 171
;   LENGTH: 561
;   TYPE: DNA
;   ORGANISM: Enterobacter cloacae
US-10-417-886-171

  Query Match           63.6%;  Score 290.2;  DB 51;  Length 561;
  Best Local Similarity 77.4%;  Pred. No. 2.5e-75;
  Matches 352;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||||||||  ||| |||||||||||| ||| ||||||||||||||||| ||| | |
Db        106 ATGAAAAACAAAACGTTGTTTATGATGTTTACATTACTGGGTGCGCCTGGGTTTGTAATC 165

Qy         61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAG 120
              ||||  | ||  |||||||| | ||| |||||||| || ||| |||||||||| |||||
Db        166 GCAGGTGATTCAGATTTAGCCAGTTCTGAATATAATTTTGCGATAAATGAATTAAGTAAA 225

Qy        121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTA 180
              ||||||| |||||||||||||| |||||||||   ||  |    |||| | || |  ||
Db        226 GCTTCATACAATCAGGCAGCCATTATTGGTCAACAGGGTTCAGGAAATAATTCTGATGTA 285

Qy        181 CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCA 240
              || |||| || || ||| |  || | ||| || | || || ||  | |  ||||| ||
Db        286 CGCCAGGACGGTTCTAAATTGCTGTCCGTTATTTCTCAGGAGGGCGGGAATAACCGCGCG 345

Qy        241 AAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAAC 300
              ||  |||| ||| ||||     |||||||||| ||||| |||||| | ||||  | ||||
Db        346 AATGTTGATCAGTCAGGGACGTATAACCTTGCTTATATCGATCAGACCGGCAACGGCAAC 405

Qy        301 GATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCT 360
              ||||| ||||||  ||| || |||| ||| || || || ||||||||||||||||| ||
Db        406 GATGCGAGTATTAAGCAGGGCGCTTTTGGCAACACCGCCATGATTATCCAGAAAGGCTCG 465

Qy        361 GGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGA 420
              ||||| |  || |||||  || |||||||||||  || ||||| ||| ||||||| ||||||
Db        466 GGTAACAGGGCGAATATAACGCAGTATGGTACGCAGAAAACAGCAGTTGTAGTACAGAGA 525

Qy        421 CAGTCGCAAATGGCTATTCGCGTGACACAACGTTA 455
              |||||||||||||||||||||| | ||||||||
Db        526 CAGTCGCAAATGGCTATTCGCGTTATTCAACGTTA 560


RESULT 6
US-10-146-492B-80
; Sequence 80, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 80
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 39-58
US-10-146-492B-80

  Query Match           21.9%;  Score 100;  DB 45;  Length 100;
  Best Local Similarity 100.0%;  Pred. No. 6.7e-19;
  Matches 100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        266 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTT 60

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 365
              |||||||||||||||||||||||||||||||||||||||
Db         61 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAA 100


RESULT 7
US-10-146-492B-81
; Sequence 81, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 81
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 38-57
US-10-146-492B-81

  Query Match           21.9%;  Score 100;  DB 45;  Length 100;
  Best Local Similarity 100.0%;  Pred. No. 6.7e-19;
  Matches 100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        208 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAAC 60

Qy        268 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 307
              |||||||||||||||||||||||||||||||||||||||
Db         61 CTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCA 100


RESULT 8
US-10-146-492B-82
; Sequence 82, Application US/10146492B
; GENERAL INFORMATION:
;  APPLICANT: MWG Biotech AG
;  TITLE OF INVENTION: Biochip
;  FILE REFERENCE: REN-14043
;  CURRENT APPLICATION NUMBER: US/10/146,492B
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: EP 0111279.5
;  PRIOR FILING DATE: 2001-05-17
;  NUMBER OF SEQ ID NOS: 12816
; SEQ ID NO 82
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Escherichia coli K-12 MG1655
;   FEATURE:
;   OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
;  PUBLICATION INFORMATION:
;   RELEVANT RESIDUES: 57-76
US-10-146-492B-82

  Query Match           21.9%;  Score 100;  DB 45;  Length 100;
  Best Local Similarity 100.0%;  Pred. No. 6.7e-19;
  Matches 100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        123 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACG 60

Qy        183 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 222
              |||||||||||||||||||||||||||||||||||||||
Db         61 GCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAA 100


RESULT 9
US-08-466-194-14/c
; Sequence 14, Application US/08466194
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, NW, Suite 500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20007
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,194
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/232,463
;      FILING DATE:  22-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/935,313
;      FILING DATE:  26-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/201 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-466-194-14

  Query Match             8.7%;  Score 39.6;  DB 8;  Length 7218;
  Best Local Similarity   10.5%;  Pred. No. 2.8;
  Matches   39;  Conservative  166;  Mismatches  165;  Indels    0;  Gaps    0;

Qy         64 GCAGGTTATGATTTAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCT 123
              ||| | | | ||   ||  | | || ||         |||| |  |   || |
Db       1523 GCATCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTACCTATCTATGCAA 1464

Qy        124 TCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGG 183
              | ||||  ||  ||    | |||||  ::  ::::  :: :: :: :   ::  : ::
Db       1463 GTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1404

Qy        184 CAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAG 243
              :::::::   ::::     :: ::  :  : : ::::::: :: :: ::  ::: ::::
Db       1403 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1344

Qy        244 ATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGAT 303
              :  ::  ::: ::::::  : ::    : : : :  ::  ::: ::: :: :  :: ::
Db       1343 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1284

Qy        304 GCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGT 363
              :  :: :    : :::: :   : :: :: :  : :: ::  :   :::::::    ::
Db       1283 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1224

Qy        364 AATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAG 423
              :: :::: ::: :  : : :: : :: :   :::::: :: ::  : :: : ::::: ::
Db       1223 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1164

Qy        424 TCGCAAATGG 433
              : ::: ::
Db       1163 RRRRRRRRRR 1154


RESULT 10
US-09-332-782-15700
; Sequence 15700, Application US/09332782
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-754CON1
;  CURRENT APPLICATION NUMBER: US/09/332,782
;  CURRENT FILING DATE: 1999-06-14
;  EARLIER APPLICATION NUMBER: US 09/181,430
;  EARLIER FILING DATE: 1998-10-28
;  NUMBER OF SEQ ID NOS: 21027
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15700
;   LENGTH: 391
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-332-782-15700

  Query Match             8.4%;  Score 38.2;  DB 18;  Length 391;
  Best Local Similarity   58.3%;  Pred. No. 2.7;
  Matches   67;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        309 TATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAA 368
              | |  |||||   |||  |||| || |||   ||  ||| | | ||   ||||| |||
Db        277 TTTCGCGCAACAAGCTGGTGGTCATGCTGAACTGGATATTCCGCAATAATCTGGAAATGT 336

Qy        369 AGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAG 423
              ||| || | | || | |||| |    | ||   | || || |   ||| | || |
Db        337 AGCCAAGACTGCAAATTATGATGACAATAATCAGACACTTCTGTAGCATATACTG 391


RESULT 11
US-09-515-694-15700
; Sequence 15700, Application US/09515694
; GENERAL INFORMATION:
;  APPLICANT: Arterburn, Matthew
;  APPLICANT: Asghari, Vida
;  APPLICANT: Blun, Linda
;  APPLICANT: Cheung, Patrick
;  APPLICANT: Damavandi, Simin
;  APPLICANT: Dickson, Mark
;  APPLICANT: Drake, Jim
;  APPLICANT: Drmanac, Radoje
;  APPLICANT: Engleman, Carrie
;  APPLICANT: Faulkner, Brandy
;  APPLICANT: Fox, Melvin
;  APPLICANT: Garcia, Veronica
;  APPLICANT: Giedt, Gretchen
;  APPLICANT: Jomek, Leni
;  APPLICANT: Jones, Lee
;  APPLICANT: Kita, David
```

```
;  APPLICANT: Kofler, Janette
;  APPLICANT: Labat, Ivan
;  APPLICANT: Lee, Won-Jae
;  APPLICANT: Lomelli, Michelle
;  APPLICANT: Nguyen, Hong
;  APPLICANT: Nguyen, Linh
;  APPLICANT: Nguyen, Lynne
;  APPLICANT: Nguyen, Phuong
;  APPLICANT: Nogra, Margie
;  APPLICANT: Ojeda, Jesse
;  APPLICANT: Palencia, Servando
;  APPLICANT: Raisi, Fariba
;  APPLICANT: Randhwa, Gurpreet
;  APPLICANT: Sahourieh, Hannah
;  APPLICANT: Sidhu, Navjiwan
;  APPLICANT: Smith, Benjamin
;  APPLICANT: Smythe, Ashleigh
;  APPLICANT: Tkach, Joe
;  APPLICANT: Tulpule, Mukul
;  APPLICANT: Verna, Ron
;  APPLICANT: Wachter, Adam
;  APPLICANT: Wu, James
;  APPLICANT: Yim, Kenneth
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 754CIP
;  CURRENT APPLICATION NUMBER: US/09/515,694
;  CURRENT FILING DATE: 2000-02-29
;  EARLIER APPLICATION NUMBER: 09/332,782
;  EARLIER FILING DATE: 1999-06-14
;  EARLIER APPLICATION NUMBER: 09/181,430
;  EARLIER FILING DATE: 1998-10-28
;  NUMBER OF SEQ ID NOS: 21027
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15700
;   LENGTH: 391
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-515-694-15700

  Query Match             8.4%;  Score 38.2;  DB 21;  Length 391;
  Best Local Similarity   58.3%;  Pred. No. 2.7;
  Matches   67;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        309 TATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAA 368
              | |  |||||   |||  |||| || |||   ||  ||| | | ||   ||||| |||
Db        277 TTTCGCGCAACAAGCTGGTGGTCATGCTGAACTGGATATTCCGCAATAATCTGGAAATGT 336

Qy        369 AGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAG 423
              ||| || | | || | |||| |     | ||   | || || |   ||| | || |
Db        337 AGCCAAGACTGCAAATTATGATGACAATAATCAGACACTTCTGTAGCATATACTG 391


RESULT 12
US-10-431-652-3221
; Sequence 3221, Application US/10431652
; GENERAL INFORMATION:
;  APPLICANT: Breton, Gary L.
;  APPLICANT:  Bush, David
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION:  BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH03-08
;  CURRENT APPLICATION NUMBER: US/10/431,652
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: US 09/328,352
;  PRIOR FILING DATE: 1999-06-04
;  PRIOR APPLICATION NUMBER: US 60/088,701
;  PRIOR FILING DATE: 1998-06-09
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3221
;   LENGTH: 4860
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-10-431-652-3221

  Query Match             8.4%;  Score 38.2;  DB 51;  Length 4860;
  Best Local Similarity   45.9%;  Pred. No. 6.3;
  Matches  130;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy        106 AATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAAT 165
              |||     |  |||| |  | |  || | | |||||     | |  |||||| |   || |
Db       3445 AATACTATTGGTAAATTATTAGGTACTGATGCAGCAGATAATCTTCAAGCAGATGCTGAG 3504

Qy        166 AATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGT 225
              ||   ||  |   |   |||||||     |       |     | || ||  ||  | | |
Db       3505 ATTTCAACTATTTATGGTTTGGGAGGCAATGATACGATTCAAGGTGGTGTACAGAATGAT 3564

Qy        226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
                |    | | |    |  ||| | |||  || || |||    |  | |  | || |
Db       3565 TATCTGTATGGTGATGATGGAGACGATACATTAGTTTCTAATACAGGCTCTGATTATTTG 3624

Qy        286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
                ||  ||||     ||| ||    |     | || | | ||   |||      | | |
Db       3625 TATGGTGGTGCAGGTAATGACACATTGATTTATGGCGGTAATTCCAATGTGTATACGGCT 3684

Qy        346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATG 388
              | ||      |||| ||| |  |  |||||| | |  | |
Db       3685 TTACAAGGTCAAGCTGGGAATGACACTTATATTATAGATAAAG 3727


RESULT 13
US-09-887-272A-5263/c
; Sequence 5263, Application US/09887272A
; GENERAL INFORMATION:
;  APPLICANT: Hou, Yu-Ming
;  APPLICANT:  Quan, Sheng
;  APPLICANT:  Chang, Hur-Song
;  APPLICANT:  Zhu, Tong
;  APPLICANT:  Whitham, Steve
;  APPLICANT:  Goff, Steve
;  APPLICANT:  Glazebrook, Jane
;  APPLICANT:  Chen, Wenquiong
;  APPLICANT:  Katagiri, Fumiaki
;  APPLICANT:  Xie, Zhiyi
;  APPLICANT:  Tao, Yi
;  APPLICANT:  Zou, Guangzhou
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: PLANT GENES INVOLVED IN DEFENSE AGAINST
;  TITLE OF INVENTION:  PATHOGENS
;  FILE REFERENCE: 1360.003US2
;  CURRENT APPLICATION NUMBER: US/09/887,272A
;  CURRENT FILING DATE:  2001-06-23
;  PRIOR APPLICATION NUMBER: 60/213,634
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/214,926
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/261,320
;  PRIOR FILING DATE: 2001-01-12
;  PRIOR APPLICATION NUMBER: 60/264,353
;  PRIOR FILING DATE: 2001-01-26
;  PRIOR APPLICATION NUMBER: 60/273,879
;  PRIOR FILING DATE: 2001-03-07
;  PRIOR APPLICATION NUMBER: 09/887,271
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 6813
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5263
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-887-272A-5263

  Query Match             8.3%;  Score 37.8;  DB 36;  Length 2000;
  Best Local Similarity   13.1%;  Pred. No. 6.2;
  Matches   36;  Conservative  121;  Mismatches  118;  Indels    0;  Gaps    0;
```

```
Qy          21 TATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATGATTTAGC 80
               :| |:::: :|  :: |  : ::: :|   : :      : ::  |:::  : :   :
Db         950 WAAGRWRWRMAWCWYCCMWKWWKMTSCMWWKYWRTWSCWYTMWWGAMRYAYYAMRRRRWT 891

Qy          81 TAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGC 140
               : : : :  :|:| : :|      :| : :||  | :::  :| : : :|:|||:   ::
Db         890 YKWSWRRMYWTMTKWAWTWMTCMCMAKWYMATGWATWMWWRYTMYTYCYAMTCAKCKYKM 831

Qy         141 CATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACT 200
               :| ::|   ||:: |: ::  :| ::: ::::: | :::  :|:: |: :     : : :
Db         830 AMTKWWTTWACAWRATSWRWRAMAGMRWKRYKMKRAYWWRWWRCWKAGWARWMKSRYRWK 771

Qy         201 TTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGA 260
               :::   : ::: : :  :::::: :   : : :   : :|::       |    :|::|::
Db         770 WKKYATRYYWKMWAMTWWWSWRRWKSYRMWSGMGRMRWSAWRYCSRMKCAKTKYASSARW 711

Qy         261 TTATAACCTTGCATATATTGATCAGGCGGGCAGTG 295
               |::  ::: :: :: :  :|:|:: :: ::|: |:
Db         710 TKRAKRSYRYRRRWYWKRKGWTYRYRYWRSCRMTR 676


RESULT 14
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             8.2%;  Score 37.4;  DB 23;  Length 456;
  Best Local Similarity   51.5%;  Pred. No. 5;
  Matches   86;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         277 ATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACT 336
               |||   |||   ||   ||  ||||  |||  | |    || |||||  ||  || ||||
Db         208 ATTACCCAGAGCGGTTATGGTAACGGCGCCGATGTAGGCCAGGGTGCGGATAATAGTACT 267

Qy         337 GCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAA 396
               | |  ||||| ||||   | ||||| || |  ||    ||||      || ||
Db         268 ATTGAACTGACTCAGAATGGTTTCAGAAATAATGCCACCATCGACCAGTGGAACGCTAAA 327

Qy         397 AAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGGCTATTCGCGT 443
               ||  | |  |||   ||       | |    ||  ||| ||  || ||
Db         328 AACTCCGATATTACTGTCGGCCAATATGATCAGCTGGTTACCCGTGT 374


RESULT 15
US-09-949-016-17017/c
; Sequence 17017, Application US/09949016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17017
;   LENGTH: 68667
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(68667)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17017

  Query Match             8.2%;  Score 37.4;  DB 39;  Length 68667;
  Best Local Similarity   58.6%;  Pred. No. 27;
  Matches   65;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386
               || | || ||| ||| | | |||  ||||  |  || ||||||  | | || | | | ||
Db       17497 TGCTCATTCTGTGATAAATTTCCTTAAAGTGTAAGGGAATAAATAATAAATAAAATATTA 17438

Qy         387 TGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGGCTAT 437
               || ||    || | | |||  |||  ||    |  | |  |||||  || |
Db       17437 TGATATGAGAATAGCAGCATGTGTGCTGTGATGTTACTTCCAAATTACTTT 17387


Search completed: March 19, 2004, 03:01:45
Job time : 2654.77 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 54.1445 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	37.4	8.2	65558	6	US-10-765-790-103	Sequence 103, App
	2	35	7.7	675	5	US-09-741-873C-3	Sequence 3, Appli
c	3	34.4	7.5	519599	6	US-10-765-790-73	Sequence 73, Appl
	4	34.2	7.5	887	6	US-10-767-701-9911	Sequence 9911, Ap
	5	33.8	7.4	1149	6	US-10-771-241-139	Sequence 139, App
c	6	33.4	7.3	38719	6	US-10-765-790-43	Sequence 43, Appl
c	7	32.2	7.1	261922	6	US-10-767-471-10653	Sequence 10653, A
	8	31.6	6.9	512	6	US-10-767-701-31198	Sequence 31198, A
	9	31.4	6.9	525	6	US-10-417-884A-2325	Sequence 2325, Ap
	10	31.4	6.9	1245	7	US-60-546-745-36	Sequence 36, Appl
	11	31.4	6.9	53332	6	US-10-786-065-3	Sequence 3, Appli
c	12	30.8	6.8	201	6	US-10-767-471-8836	Sequence 8836, Ap
c	13	30.8	6.8	201	6	US-10-767-471-34493	Sequence 34493, A
c	14	30.8	6.8	1227	6	US-10-767-471-501	Sequence 501, App
c	15	30.8	6.8	76573	6	US-10-767-471-10718	Sequence 10718, A
c	16	30.4	6.7	1227	6	US-10-417-884A-2544	Sequence 2544, Ap
	17	30.2	6.6	3529	6	US-10-779-543-9918	Sequence 9918, Ap
c	18	30.2	6.6	1790242	6	US-10-767-471-10805	Sequence 10805, A
c	19	29.8	6.5	600	7	US-60-545-213-3515	Sequence 3515, Ap
c	20	29.8	6.5	600	7	US-60-545-213-7787	Sequence 7787, Ap
c	21	29.6	6.5	201	6	US-10-767-471-8837	Sequence 8837, Ap
c	22	29.6	6.5	201	6	US-10-767-471-34439	Sequence 34439, A
	23	29.6	6.5	40000	6	US-10-767-471-10911	Sequence 10911, A
c	24	29.6	6.5	122673	6	US-10-765-790-33	Sequence 33, Appl
	25	29.4	6.4	2806	6	US-10-021-698A-2111	Sequence 2111, Ap
	26	29.4	6.4	74677	6	US-10-767-471-10891	Sequence 10891, A
	27	29.2	6.4	1080	6	US-10-451-467A-619	Sequence 619, App
c	28	29.2	6.4	134134	6	US-10-767-471-10743	Sequence 10743, A
c	29	29.2	6.4	189268	6	US-10-021-698A-697	Sequence 697, App
	30	29.2	6.4	290892	6	US-10-767-471-10809	Sequence 10809, A
c	31	29	6.4	447	6	US-10-767-701-18004	Sequence 18004, A
	32	29	6.4	532	6	US-10-767-701-17618	Sequence 17618, A
c	33	29	6.4	13948	6	US-10-767-471-10858	Sequence 10858, A
	34	29	6.4	16201	6	US-10-767-471-10705	Sequence 10705, A
	35	29	6.4	45268	7	US-60-548-091-5603	Sequence 5603, Ap
c	36	28.8	6.3	1829	6	US-10-775-169-326	Sequence 326, App
	37	28.8	6.3	1914	6	US-10-767-701-13618	Sequence 13618, A
	38	28.8	6.3	20767	6	US-10-100-683-12012	Sequence 12012, A
	39	28.6	6.3	408	6	US-10-417-884A-793	Sequence 793, App
c	40	28.6	6.3	1056	6	US-10-781-014-285	Sequence 285, App
c	41	28.6	6.3	1056	6	US-10-781-014-449	Sequence 449, App
	42	28.6	6.3	78785	5	US-09-978-167A-3	Sequence 3, Appli
	43	28.6	6.3	100374	7	US-60-548-091-5675	Sequence 5675, Ap
c	44	28.6	6.3	189817	6	US-10-767-471-10675	Sequence 10675, A
c	45	28.4	6.2	477	6	US-10-779-543-18788	Sequence 18788, A

ALIGNMENTS

RESULT 1
US-10-765-790-103/c
; Sequence 103, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 103
; LENGTH: 65558
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-103

Query Match 8.2%; Score 37.4; DB 6; Length 65558;
Best Local Similarity 58.6%; Pred. No. 0.21;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy        327 TGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTA 386
              || | || ||| ||| | | |||  ||||  |  || ||||||  | | || | | | ||
Db      16503 TGCTCATTCTGTGATAAATTTCCTTAAAGTGTAAGGGAATAAATAATAAATAAAATATTA 16444

Qy        387 TGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGGCTAT 437
              || ||    || | | |||  |||  ||    |  | |  |||||  || |
Db      16443 TGATATGAGAATAGCAGCATGTGTGCTGTGATGTTACTTCCAAATTACTTT 16393
```

RESULT 2
US-09-741-873C-3
; Sequence 3, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1

```
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/978,878
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 07/347,189
;  PRIOR FILING DATE: 1989-05-04
;  PRIOR APPLICATION NUMBER: US 07/789,437
;  PRIOR FILING DATE: 1991-11-06
;  PRIOR APPLICATION NUMBER: US 07/970,846
;  PRIOR FILING DATE: 1992-11-03
;  PRIOR APPLICATION NUMBER: US 08/187,865
;  PRIOR FILING DATE: 1994-01-28
;  PRIOR APPLICATION NUMBER: US 08/318,519
;  PRIOR FILING DATE: 1994-10-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 675
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (83)..(538)
US-09-741-873C-3

  Query Match             7.7%;  Score 35;  DB 5;  Length 675;
  Best Local Similarity   100.0%;  Pred. No. 0.24;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        416 AGAGACAGTCGCAAATGGCTATTCGCGTGACACAA 450
              ||||||||||||||||||||||||||||||||||||
Db          1 AGAGACAGTCGCAAATGGCTATTCGCGTGACACAA 35


RESULT 3
US-10-765-790-73/c
; Sequence 73, Application US/10765790
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73
;   LENGTH: 519599
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-73

  Query Match             7.5%;  Score 34.4;  DB 6;  Length 519599;
  Best Local Similarity   60.9%;  Pred. No. 3.4;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        263 ATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTG 322
              ||||  ||| || |  |||||||  |    | || | |  || ||| |||| | || | |
Db     332273 ATAAAATTGAATTTTGTGATCAGTAGCCATGGGCTATCTCTGACAGCATTTGGAAAAGAG 332214

Qy        323 CTTATGGTAATACTGCGATGATTATCCAGAAA 354
              ||   | | |||| |  |  || | ||| |||
Db     332213 CTATAGTTGATACAGAAAATATAACCCAAAAA 332182


RESULT 4
US-10-767-701-9911
; Sequence 9911, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 9911
;   LENGTH: 887
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS40598_1
US-10-767-701-9911

  Query Match             7.5%;  Score 34.2;  DB 6;  Length 887;
  Best Local Similarity   53.3%;  Pred. No. 0.47;
  Matches   72;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        203 TGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATT 262
              || |  |  || |  |||  |||  ||  |   ||| |||  | ||   | |   |||
Db        433 TGTCAATGATTCCTAAAGGTGGTGATACAATTTGGGGAAATCTGGATTGGTCTCCAGAAG 492

Qy        263 ATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTG 322
              ||  ||||| || ||  | | ||  |  ||    |||| ||| ||    |||| | | |
Db        493 ATGGCCTTGAATGTAAAGCTAAGAAGCACAAAACCAATGATACCGAGGTTTCTAAGGATA 552

Qy        323 CTTATGGTAATACTG 337
              ||||  | | ||
Db        553 GCAATGGGGAAAATG 567


RESULT 5
US-10-771-241-139
; Sequence 139, Application US/10771241
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Forsyth, R. Allyn
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001C1
;  CURRENT APPLICATION NUMBER: US/10/771,241
;  CURRENT FILING DATE:  2004-02-03
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 139
;   LENGTH: 1149
;   TYPE: DNA
;   ORGANISM: E. Coli
US-10-771-241-139

  Query Match             7.4%;  Score 33.8;  DB 6;  Length 1149;
  Best Local Similarity   56.9%;  Pred. No. 0.67;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        259 GATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAA 318
              ||| |||||||    ||||  | |||||  || |   || |||  |  | ||  |||
Db        442 GATAATAACCTGAATTATAGCGTTCAGGTCGGTAACACCCACGGAGGTAATACATCGTCT 501

Qy        319 GGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATA 367
              ||  |  |||| | | | | | | |    || || ||||||||
Db        502 GGCACCAGTGGTTACAGTTCTCTTAATTATCGTGGAGCTTATGGTAATA 550


RESULT 6
```

```
US-10-765-790-43/c
; Sequence 43, Application US/10765790
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43
;   LENGTH: 38719
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-43

  Query Match             7.3%;  Score 33.4;  DB 6;  Length 38719;
  Best Local Similarity   51.7%;  Pred. No. 2.9;
  Matches   76;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         17 TATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCCGCAGCAGGTTATGATT 76
              | || ||||| |  | || |||    | | |    |  || | |       ||| |
Db      11094 TGTTGATGATATGGAGAAATCTGAACCCTCATACACTGCTAGTGGAAATGTAAAATGGTG 11035

Qy         77 TAGCTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGG 136
              |||||| || |||| |||| || || ||     || ||| ||| | |  | |||  |
Db      11034 TAGCTACTTTAGAAAATAAGTTAGCAGTTCCACAAATGATTAAATATAGAGTTACCCTAT 10975

Qy        137 CAGCCATAATTGGTCAAGCTGGGACTA 163
              | |||  | |    |  || |   ||
Db      10974 GATCCAGCAATTACAATCCTAGTGATA 10948


RESULT 7
US-10-767-471-10653/c
; Sequence 10653, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10653
;   LENGTH: 261922
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(261922)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10653

  Query Match             7.1%;  Score 32.2;  DB 6;  Length 261922;
  Best Local Similarity   51.0%;  Pred. No. 13;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        251 AGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTA 310
              |.||||    | | ||| ||| | | | |||  | ||   |  |   ||   |  || |
Db      72532 AAACAGAGCCTAAAAACATTGGAAAGAATGAATAAGCCATCTTTTAAAAAATTAACAATT 72473

Qy        311 TTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAG 370
              ||   |||| |   |   ||  |      | ||||   ||  |   || |  ||| ||||
Db      72472 TTATCCAAGTTTTATTATGTTTTTAAATTAGGATTTATCAATATCCTTTTTTTAAAAAAG 72413

Qy        371 CAAATATTACACAGTATGGTACTCAAAAA 399
              ||| | | || | ||  | | | ||| |
Db      72412 GAAACAATTCATAATAGTGGATTTAAACA 72384


RESULT 8
US-10-767-701-31198
; Sequence 31198, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 31198
;   LENGTH: 512
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(512)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 18065357
US-10-767-701-31198

  Query Match             6.9%;  Score 31.6;  DB 6;  Length 512;
  Best Local Similarity   50.7%;  Pred. No. 2.4;
  Matches   76;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        132 TCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGG 191
              | ||||   |||    | ||| | |||  | | | ||   |   ||     |||||||
Db        110 TGAGGCTCACATGGAAGTTCATGTTGGATCAATTGATGAGGAGGAGGACGAGCAGGGAAT 169

Qy        192 CTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCA 251
              || |  |  | |   ||| |||      || ||||  || || | |||| ||
Db        170 TGCACATATGATTGAACATGTCGCGTTTCTTGGCAGTAAAAAACGTGAAAAGCTTTTGGG 229

Qy        252 GACAGGAGATTATAACCTTGCATATATTGA 281
              ||| || |          ||||||||  ||
Db        230 GACCGGTGCGAGGTCTAATGCATATACAGA 259


RESULT 9
US-10-417-884A-2325
; Sequence 2325, Application US/10417884A
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/417,884A
;             FILING DATE: 17-Apr-2003
;        PRIOR APPLICATION DATA:
```

```
;             APPLICATION NUMBER: US/09/107,532
;             FILING DATE: 30-Jun-1998
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 2325:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 525 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;       FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...525
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2325:
US-10-417-884A-2325

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 525;
  Best Local Similarity   54.9%;  Pred. No. 2.8;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        270 TGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGG 329
              || ||    |||| |||||       |      | ||||| ||||||   |||   |  ||
Db        213 TGAATTATCTGATAAGGCGAAAGAAGAAGCAGTTGCCATTATTTCTTCAGGGATTAGTGA 272

Qy        330 TAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACAC 382
                |||  ||  |||| |    | |    || |||  ||||||  || |||  ||
Db        273 CTATATTGATATGAGTGGTGAAATCAATTATGGCGATAAAGACAAAATTGTAC 325


RESULT 10
US-60-546-745-36
; Sequence 36, Application US/60546745
; GENERAL INFORMATION:
;  APPLICANT: Klaenhammer, Todd R.
;  APPLICANT:  Russell, William M.
;  APPLICANT:  Alterman, Eric
;  APPLICANT:  Cano, Raul
;  APPLICANT:  Hamrick, Alice
;  TITLE OF INVENTION: LACTOBACILLUS ACIDOPHILUS NUCLEIC ACID SEQUENCES ENCODING
;  TITLE OF INVENTION:  PROTEASE HOMOLOGUES AND USES THEREOF
;  FILE REFERENCE: 5051.604PR2
;  CURRENT APPLICATION NUMBER: US/60/546,745
;  CURRENT FILING DATE:  2004-02-23
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
;   LENGTH: 1245
;   TYPE: DNA
;   ORGANISM: Lactobacillus acidophilus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1245)
;   OTHER INFORMATION: pepT amino tripeptidase T (EC 3.411.-) ORF# 1190
US-60-546-745-36

  Query Match             6.9%;  Score 31.4;  DB 7;  Length 1245;
  Best Local Similarity   49.7%;  Pred. No. 3.7;
  Matches   80;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        240 AAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAA 299
              | |  ||||  |   |||      |  |  ||   || | | |  |||| || ||||  |
Db        129 AGATCTTGAGGAATTAGGCTTGAAAGATATTCACTACAATCAAAAGGCTGGTAGTGTAAT 188

Qy        300 CGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTC 359
              || ||  | |  |||  | | || ||||| |    |||  |||    ||  |  |  |
Db        189 CGCTGAAATTCCTTCAAATGTTGATTATGATGTACCTGTAATGGGATTCTTGGCACATAG 248

Qy        360 TGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAA 400
              || ||      |     | | || |  ||| ||  | | |||
Db        249 TGATACAGCTGATTTCAATTCAGAAAATGTTAAACCACAAA 289


RESULT 11
US-10-786-065-3
; Sequence 3, Application US/10786065
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001098DIV II
;  CURRENT APPLICATION NUMBER: US/10/786,065
;  CURRENT FILING DATE:  2004-02-26
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 53332
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-786-065-3

  Query Match             6.9%;  Score 31.4;  DB 6;  Length 53332;
  Best Local Similarity   61.7%;  Pred. No. 13;
  Matches   50;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy        336 TGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCA 395
              ||  |||| ||| ||  |||  | ||   |||||||  ||| ||| |||  | |  |||
Db       8632 TGTTATGAGTATTCAACAAGAATATGTATATAAAGCACTTATCACAGAGTCAGTTTTTCA 8691

Qy        396 AAAAACGGCAATTGTAGTGCA 416
              |   | ||| ||| |  ||
Db       8692 GAGTTCAACAAATGTTGACCA 8712


RESULT 12
US-10-767-471-8836/c
; Sequence 8836, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8836
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-8836

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 201;
  Best Local Similarity   54.6%;  Pred. No. 3.1;
  Matches   59;  Conservative    1;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGT 385
              |||| | ||  | |  ||  |||||  | |||  || ||||:| |    |||| | |||
Db        142 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 83

Qy        386 ATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGG 433
              | | |  |||   |   |     | || ||||  ||||    |  |||
```

```
Db         82 AGGATTATCATGTACATGATGAGGCAGAGCAGGAACAGAAAAAGCTGG 35


RESULT 13
US-10-767-471-34493/c
; Sequence 34493, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34493
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-34493

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 201;
  Best Local Similarity   54.6%;  Pred. No. 3.1;
  Matches   59;  Conservative    1;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGT 385
              |||| | ||  | |  ||  |||||  | |||  || ||||:| |    |||| | |||
Db        142 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 83

Qy        386 ATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGG 433
              | | |  |||   |   |     | || ||||  ||||    |  |||
Db         82 AGGATTATCATGTACATGATGAGGCAGAGCAGGAACAGAAAAAGCTGG 35


RESULT 14
US-10-767-471-501/c
; Sequence 501, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 501
;   LENGTH: 1227
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-501

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 1227;
  Best Local Similarity   54.6%;  Pred. No. 5.7;
  Matches   59;  Conservative    1;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGT 385
              |||| | ||  | |  ||  |||||  | |||  || ||||:| |    |||| | |||
Db        476 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 417

Qy        386 ATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGG 433
              | | |  |||   |   |     | || ||||  ||||    |  |||
Db        416 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 369


RESULT 15
US-10-767-471-10718/c
; Sequence 10718, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10718
;   LENGTH: 76573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(76573)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-767-471-10718

  Query Match             6.8%;  Score 30.8;  DB 6;  Length 76573;
  Best Local Similarity   54.6%;  Pred. No. 22;
  Matches   59;  Conservative    1;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        326 ATGGTAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGT 385
              |||| | ||  | |  ||  |||||  | |||  || ||||:| |    |||| | |||
Db      69821 ATGGGAGTATGGAGGCGAGAATCCAAGATGGTGATGATAATRAGGAGGCTATTTCCCAGG 69762

Qy        386 ATGGTACTCAAAAAACGGCAATTGTAGTGCAGAGACAGTCGCAAATGG 433
              | | |  |||   |   |     | || ||||  ||||    |  |||
Db      69761 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGG 69714


Search completed: March 19, 2004, 03:10:03
Job time : 57.1445 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 2033.55 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaaaacaaattgttatt.....ttcgcgtgacacaacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	205.8	45.1	220	12	BI941542	BI941542 dg26g06.y
c	2	56.8	12.5	207	9	AI368352	AI368352 SWOvMfCAR
c	3	47	10.3	680	12	BJ618688	BJ618688 BJ618688
c	4	40.2	8.8	956	29	CNS04QM8	AL302777 Tetraodon
	5	37.8	8.3	533	10	BB703456	BB703456 BB703456
	6	37	8.1	562	14	CF068694	CF068694 EST669415
	7	37	8.1	671	12	BI311207	BI311207 EST531295
	8	37	8.1	756	12	BM815787	BM815787 EST593881
c	9	36.6	8.0	957	13	BX453027	BX453027 BX453027
	10	36.4	8.0	364	12	BM540455	BM540455 qj59d01.b
	11	36.4	8.0	597	28	AZ169114	AZ169114 SP_0113_A
	12	36.4	8.0	657	10	BB449135	BB449135 BB449135
	13	36.4	8.0	769	12	BG354883	BG354883 MBTsMLA13
	14	36.2	7.9	491	28	AZ811111	AZ811111 2M0077A11
	15	36.2	7.9	532	28	AZ811109	AZ811109 2M0077A09
	16	36	7.9	678	12	BI452381	BI452381 GI09_G10
	17	36	7.9	818	9	AL716706	AL716706 AL716706
	18	35.6	7.8	615	28	AQ308392	AQ308392 CITBI-E1-
	19	35.6	7.8	624	28	AQ779365	AQ779365 HS_3207_B
	20	35.4	7.8	705	12	BG125943	BG125943 EST471589
	21	35.2	7.7	563	14	CA855727	CA855727 PfESToac3
c	22	35.2	7.7	744	12	BJ433259	BJ433259 BJ433259
c	23	35.2	7.7	1101	29	CNS0039G	AL063921 Drosophil
c	24	35.2	7.7	1134	9	AL551649	AL551649 AL551649
c	25	35	7.7	491	28	BH068121	BH068121 RPCI-24-3
c	26	34.8	7.6	774	28	BZ527786	BZ527786 OGAJU41TC
	27	34.8	7.6	788	14	CA766697	CA766697 AF53-Rpf_
	28	34.8	7.6	823	28	BZ527794	BZ527794 OGAJU41TM
	29	34.8	7.6	992	13	BU180244	BU180244 AGENCOURT
c	30	34.6	7.6	388	28	B86074	B86074 RPCI11-21O1
	31	34.6	7.6	572	13	BX552747	BX552747 BX552747
	32	34.6	7.6	1200	13	BX399167	BX399167 BX399167
	33	34.4	7.5	412	12	BM167028	BM167028 EST569551
c	34	34.4	7.5	413	9	AL698409	AL698409 DKFZp686I
	35	34.4	7.5	595	12	BM852853	BM852853 K-EST0134
c	36	34.4	7.5	849	29	CG950661	CG950661 MBEHO26TF
c	37	34.4	7.5	862	29	CNS029ER	AL187164 Tetraodon
c	38	34.4	7.5	876	28	BZ794472	BZ794472 PUFFR15TD
c	39	34.4	7.5	1050	13	BX349695	BX349695 BX349695
c	40	34.4	7.5	1201	13	BX324630	BX324630 BX324630
c	41	34.4	7.5	4397	29	AY413040	AY413040 Homo sapi
	42	34.2	7.5	528	10	BE599976	BE599976 PI1_78_A0
c	43	34.2	7.5	528	28	AQ606604	AQ606604 HS_2095_B
	44	34.2	7.5	531	12	BM323404	BM323404 PIC1_18_H
	45	34.2	7.5	623	29	CC816176	CC816176 1O0001A10

ALIGNMENTS

RESULT 1
BI941542
LOCUS BI941542 220 bp mRNA linear EST 19-FEB-2003
DEFINITION dg26g06.y1 Xenopus laevis gastrula non normalized Xenopus laevis cDNA clone IMAGE:3749963 5' similar to SW:CSGB_ECOLI P39828 MINOR CURLIN SUBUNIT PRECURSOR. ;, mRNA sequence.
ACCESSION BI941542
VERSION BI941542.1 GI:16256014
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 220)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Source lab clone id - 3749963
Trace considered overall poor quality This clone is available royalty-free through LLNL; contact the IMAGE Consortium (image@image.llnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .220
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:3749963"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F'"
/clone_lib="Xenopus laevis gastrula non normalized"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2 ug of poly A+ RNA (equal parts from stage 10.5 and stage 11.5 gastrulae). EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. The library was mass excised and used to infect Top10F'. Clones were picked into freezing medium (per liter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM K2HPO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4 7 H2O, 6.8 mM (NH4)2SO4, 4 % w/v glycerol) and grown for 24 hours. Original library construction by Bruce Blumberg (Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

Query Match 45.1%; Score 205.8; DB 12; Length 220;
Best Local Similarity 96.8%; Pred. No. 4.7e-51;
Matches 210; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy         80 CTAATTCAGAATATAACTTCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAG 139
              |||||||| |||||||||||||| | |||||| ||||||||||||||||||||||||||
Db          3 CTAATTCATAATATAACTTCGCGGCACATGAATCGAGTAAGTCTTCATTTAATCAGGCAG 62

Qy        140 CCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAAC 199
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db         63 CCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACTGCAGGGAGGCTCAAAAC 122

Qy        200 TTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAG 259
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TTTTGGCGGTTGCTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAG 182

Qy        260 ATTATAACCTTGCATATATTGATCAGGCGGGCAGTGC 296
              |||||||||||||||||||||||||||||||||||| |
Db        183 ATTATAACCTTGCATATATTGATCAGGCGGGCAGTCC 219
```

RESULT 2
AI368352/c
LOCUS AI368352 207 bp mRNA linear EST 11-JAN-1999
DEFINITION SWOvMfCAR04G09SK Onchocerca volvulus microfilaria cDNA (SAW98MLW-OvMf) Onchocerca volvulus cDNA clone SWOvMfCAR04G09 5', mRNA sequence.
ACCESSION AI368352
VERSION AI368352.1 GI:4147105
KEYWORDS EST.
SOURCE Onchocerca volvulus
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 207)
AUTHORS Williams,S.A.
TITLE Genes expressed in microfilaria of Onchocerca volvulus
JOURNAL Unpublished (1999)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.

FEATURES Location/Qualifiers
source 1. .207
/organism="Onchocerca volvulus"
/mol_type="mRNA"
/db_xref="taxon:6282"
/clone="SWOvMfCAR04G09"
/dev_stage="microfilaria"
/lab_host="XL1-Blue MRF'"
/clone_lib="Onchocerca volvulus microfilaria cDNA (SAW98MLW-OvMf)"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.8 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email:genome@smith.edu."

ORIGIN

Query Match 12.5%; Score 56.8; DB 9; Length 207;
Best Local Similarity 95.1%; Pred. No. 3.9e-06;
Matches 58; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATGAAAAACAAATTGTTATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 60
              ||||||| | |||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGAAAACCNAATTGTCATTTATGATGTTAACAATACTGGGTGCGCCTGGGATTGCAGCC 2

Qy         61 G 61
              |
Db          1 G 1
```

RESULT 3
BJ618688/c
LOCUS BJ618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BJ618688 NIBB Mochii normalized Xenopus early gastrula library Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BJ618688
VERSION BJ618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.
http://xenopus.nibb.ac.jp.

FEATURES Location/Qualifiers
source 1. .680

```
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL186b22"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
ORIGIN

  Query Match             10.3%;  Score 47;  DB 12;  Length 680;
  Best Local Similarity   100.0%;  Pred. No. 0.005;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        410 TAGTGCAGAGACAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 456
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        680 TAGTGCAGAGACAGTCGCAAATGGCTATTCGCGTGACACAACGTTAA 634


RESULT 4
CNS04QM8/c
LOCUS       CNS04QM8                 956 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            129D18 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL302777
VERSION     AL302777.1  GI:8183119
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 956)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="129D18"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG129DB09SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             8.8%;  Score 40.2;  DB 29;  Length 956;
  Best Local Similarity   46.6%;  Pred. No. 0.62;
  Matches  129;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        113 TGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTG 172
              | ||| ||||   |   | | |   | |  | ||  ||||  | |||   | ||||
Db        430 TCAGTCAGTCAGGAGCCACTAATCAATCACTGATCAGTCAGTCAGGGGTCACTAATCAAT 371

Qy        173 CTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCA 232
              | | | |  | |||  |||  | |   |   |  | |   ||  |||   |  |
Db        370 CACTGATCAGTCAGTCAGGAGCCACTAATCAATCACTGATCAGTCAGTCAGGGGTCACTA 311

Qy        233 ACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCA 292
              | |   ||  |||    ||| ||||||    ||| |   ||   ||   |||| | ||
Db        310 ATCAATCACTGATCAGTCAGTCAGGAGTCACTAATCAATCACTGATCAGTCAGTCAGGGG 251

Qy        293 GTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGA 352
              | ||  |  |    ||    ||    |       ||||    |  |||| |  |||
Db        250 TCACTAATCAATCACTGATCAGTCAGTCAGGGGTCACTAATCAATCACTGATCAGTCAGT 191

Qy        353 AAGGTTCTGGTAATAAAGCAAATATTACACAGTATGG 389
              |||  |  |||| || ||  || | ||||  ||
Db        190 CAGGAGCCACTAATCAATCACTGATCAGTCAGTCAGG 154


RESULT 5
BB703456
LOCUS       BB703456                 533 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  BB703456 RIKEN full-length enriched, in vitro fertilized eggs Mus
            musculus cDNA clone 7420452F19 3', mRNA sequence.
ACCESSION   BB703456
VERSION     BB703456.1  GI:16052291
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
            Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
```

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .533
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="7420452F19"
/sex="female"
/tissue_type="in vitro fertilized eggs"
/dev_stage="egg"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, in vitro fertilized eggs"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match 8.3%; Score 37.8; DB 10; Length 533;
Best Local Similarity 49.3%; Pred. No. 2.8;
Matches 99; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
Qy        126 ATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCA 185
              ||||  | || | | |||||||     || |||  || |  ||| | |    ||  | |
Db        202 ATTTTTTTAGCCGGACATAATGTTGCATGCTTTGAATCCTAACATTCAGAAGCCAGAGGC 261

Qy        186 GGGAGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGAT 245
               || ||       |  | |   ||  ||| | ||  | ||    |     || ||| || | |
Db        262 AGGTGGAGATCTATATATACAGGGAGTTCCTCACTAGGCAGGGCTACACTGGGAAATGCT 321

Qy        246 TGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACGATGC 305
                ||   ||  | ||| |    ||| || |||||||  |  ||    ||   || | ||
Db        322 GCCCCCCCCAAAAAATTGTTGAGTTGTATTTATTGATTTGAGGGCTTATGTAGACCAGGC 381

Qy        306 CAGTATTTCGCAAGGTGCTTA 326
               |  ||   |  | ||  ||
Db        382 TATCCTTGAACTTGTTGTCTA 402
```

RESULT 6
CF068694
LOCUS CF068694 562 bp mRNA linear EST 21-JUL-2003
DEFINITION EST669415 MTUS Medicago truncatula cDNA clone MTUS-11C11, mRNA sequence.
ACCESSION CF068694
VERSION CF068694.1 GI:33105353
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 562)
AUTHORS VandenBosch,K., Endre,G., Silverstein,K, Town,C.D., Van Aken,S., Utterback,T., Cheung,F. and Fraser,C.M.
TITLE The Medicago truncatula 6K unigene set: cDNA clones selected and re-arrayed from various libraries
JOURNAL Unpublished (2002)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
TIGR sequence name: MTUSK35TK Alias Clone name:pGESD10A6 More information is available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .562
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="MTUS-11C11"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/clone_lib="MTUS"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 562;
Best Local Similarity 53.9%; Pred. No. 4.9;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        242 AGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACG 301
              |||||||  ||||  |  |||  ||  |  | |  ||  |  |||||       | ||   |
Db         59 AGATTGAACAGAGTGATGATAATGATGTGGATTACAAAAATCAGCTTCTTAATGTTCTCA 118

Qy        302 ATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTG 361
              |   |||  |    | |||  ||  |  ||  ||| ||||||  ||   |  |  |  | |
Db        119 AAATCAGCAAGGAGAAAGTTGATGAATGTTATAATGCGATTCTTCATCTTACAAATACAA 178

Qy        362 GTAATAAAGCAAATATTACAC 382
               ||||  |  |   |||  ||  ||
Db        179 ATAATTATGGTTATAATAAAC 199
```

RESULT 7
BI311207
LOCUS BI311207 671 bp mRNA linear EST 20-JUL-2001
DEFINITION EST5312957 GESD Medicago truncatula cDNA clone pGESD10A6 5' end, mRNA sequence.
ACCESSION BI311207
VERSION BI311207.1 GI:14985534
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 671)
AUTHORS Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from developing reproductive tissues of Medicago truncatula
JOURNAL Unpublished (2001)
COMMENT Contact: Michael A. Grusak
USDA/ARS Children's Nutrition Research Center

Baylor College of Medicine
1100 Bates Street, Houston, TX 77030-2600, USA
Tel: 713-798-7044
Fax: 713-798-7078
Email: mgrusak@bcm.tmc.edu
B398062e
TIGR sequence name: MTPAR03TK
More information is available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .671
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pGESD10A6"
/tissue_type="immature seeds"
/dev_stage="Immature seeds, 11 to 19 days after pollination"
/clone_lib="GESD"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Immature seeds, collected from pods ranging in age from 11 to 19 days after pollination, were harvested from greenhouse-grown plants. Seeds were removed and separated from pod walls and were immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propogated in XLOLR cells."

ORIGIN

Query Match 8.1%; Score 37; DB 12; Length 671;
Best Local Similarity 53.9%; Pred. No. 5.1;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        242 AGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACG 301
              ||||||| ||||  |  ||| || |  | |  || |   |||||      | ||    |
Db         79 AGATTGAACAGAGTGATGATAATGATGTGGATTACAAAAATCAGCTTCTTAATGTTCTCA 138

Qy        302 ATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTG 361
              |   ||| |    | ||| || | |  || ||| ||||||  ||   |  | |  | |
Db        139 AAATCAGCAAGGAGAAAGTTGATGAATGTTATAATGCGATTCTTCATCTTACAAATACAA 198

Qy        362 GTAATAAAGCAAATATTACAC 382
              ||||  | |   ||| || ||
Db        199 ATAATTATGGTTATAATAAAC 219
```

RESULT 8
BM815787
LOCUS BM815787 756 bp mRNA linear EST 05-MAR-2002
DEFINITION EST593881 BNIR Medicago truncatula/Meloidogyne incognita mixed EST library cDNA clone pBNIR-28N9, mRNA sequence.
ACCESSION BM815787
VERSION BM815787.1 GI:19151801
KEYWORDS EST.
SOURCE Medicago truncatula/Meloidogyne incognita mixed EST library
ORGANISM Medicago truncatula/Meloidogyne incognita mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 756)
AUTHORS Bird,D., Koltai,H., Samac,D., Town,C.D., Van Aken,S., Utterback,T., Cheung,F., Tsai,J. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula after infection with the nematode Meloidogyne incognita
JOURNAL Unpublished (2002)
COMMENT Contact: Bird, DM
Plant Nematode Genetics Group
North Carolina State University
Box 7616, Raleigh, NC 27695, USA
Tel: 919 515 6813
Fax: 919 515 9500
Email: david_bird@ncsu.edu
TIGR sequence name: MTQBI77TK More information is available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .756
/organism="Medicago truncatula/Meloidogyne incognita mixed EST library"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:188702"
/clone="pBNIR-28N9"
/tissue_type="3 week old roots"
/dev_stage="3 days after infection with Meloidogyne incognita second stage larvae"
/lab_host="XLOLR"
/clone_lib="BNIR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells. "

ORIGIN

Query Match 8.1%; Score 37; DB 12; Length 756;
Best Local Similarity 53.9%; Pred. No. 5.3;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy        242 AGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCAACG 301
              ||||||| ||||  |  ||| || |  | |  || |   |||||      | ||    |
Db        251 AGATTGAACAGAGTGATGATAATGATGTGGATTACAAAAATCAGCTTCTTAATGTTCTCA 310

Qy        302 ATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTTCTG 361
              |   ||| |    | ||| || | |  || ||| ||||||  ||   |  | |  | |
Db        311 AAATCAGCAAGGAGAAAGTTGATGAATGTTATAATGCGATTCTTCATCTTACAAATACAA 370

Qy        362 GTAATAAAGCAAATATTACAC 382
              ||||  | |   ||| || ||
Db        371 ATAATTATGGTTATAATAAAC 391
```

RESULT 9
BX453027/c
LOCUS BX453027 957 bp mRNA linear EST 22-MAY-2003
DEFINITION BX453027 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF012YJ13 3-PRIME, mRNA sequence.
ACCESSION BX453027
VERSION BX453027.1 GI:31036338
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE . Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
was not normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6669.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAH007ZA11NP1&cluster=6669.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0BAH007ZA11NP1.

```
FEATURES             Location/Qualifiers
     source          1. .957
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF012YJ13"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             8.0%;  Score 36.6;  DB 13;  Length 957;
  Best Local Similarity   32.1%;  Pred. No. 7.5;
  Matches  106;  Conservative   55;  Mismatches  169;  Indels    0;  Gaps    0;

Qy         98 TCGCGGTAAATGAATTGAGTAAGTCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTG 157
              | | ||  |  | | ::| |||: ::    : ||: |::     | ||: |   | | :|
Db        920 TGGGGGAGANGGGARWKATTAAKAVKGGGGDAAAKGAKKWTAAAAAAAKGGAGAAGGTKG 861

Qy        158 GGACTAATAATAGTGCTCAGTTACGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGCGC 217
              ||     : |  : |  :|   |:|| | :   :   |||:    ||| |      |:
Db        860 GGTTATTKGASSSSGGRSAAAAASGGKAWKAKTKAGAAAAWKAGGTGGTGAAAAATKKTA 801

Qy        218 AAGAAGGTAGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATA 277
                : ||  |::  ||    : |  :    |  |:|    || |  :||  ||:|   |
Db        800 KTAWKGGGKGKRAAAAWKAASAAKTDGGAMAAAARATKAAAGTTATKAAGWATGVAAKAA 741

Qy        278 TTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTG 337
              :  |  |      : :::   ||  |:|  |    :  : ||:        ||  |:| :|
Db        740 KGAAAAAAAKAAKGWKKAGAAAAAAKGGAAAAGGKAAKAAAKAAAAAAKGGGGKAKAGKG 681

Qy        338 CGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAA 397
                |  |::|   | ||| |:   :| |:||||: ||| |     | | : :||:     |||
Db        680 GAAGAAWKAATGAAAAAAGKAAAKGAAWTAAAKGAAAAAAAGAAGAAKGKGGKKGGAAAA 621

Qy        398 AAACGGCAATTGTAGTGCAGAGACAGTCGC 427
               ||:|| :||:|: | |    |  |: :::
Db        620 TAASGGAWATKGKGGAGKSAKGGAASASSS 591


RESULT 10
BM540455
LOCUS       BM540455                364 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  qj59d01.b1 Canis cDNAs from testes cells Canis familiaris cDNA
            clone qj59d01 5', mRNA sequence.
ACCESSION   BM540455
VERSION     BM540455.1  GI:18823255
KEYWORDS    EST.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   O'Shaughnessy,A.L., McCombie,W.R., Baker,J.P., Balija,V.,
            Cunnius,D., Dedhia,N.N., de la Bastide,M., Katzenberger,F.,
            King,L., Kirchoff,K.A., Miller,B., Muller,S., Nascimento,L.U.,
            Palmer,L., Santos,L., Shah,R.S., Spiegel,L.A., Zutavern,T.,
            Preston,R. and Hannon,G.J.
  TITLE     Expressed sequence tags from Canis familiaris (dog) (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: qj59  row: d  column: 01
            Seq primer: -21M13UnivFwd
            High quality sequence stop: 364.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Canis familiaris"
                     /mol_type="mRNA"
                     /db_xref="taxon:9615"
                     /clone="qj59d01"
                     /tissue_type="testes"
                     /clone_lib="Canis cDNAs from testes cells"
                     /note="Vector: Lambda Zap II; The library was produced by
                     Greg Hannon and Raymond Preston (Cold Spring  Harbor
                     Laboratory). This library is oligo(dT) primed using
                     stratagene  zap cDNA synthesis kit. It was made from dog
                     testes. Please contact Greg  Hannon (hannon@cshl.org) with
                     any library related inquiries."
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 12;  Length 364;
  Best Local Similarity   49.0%;  Pred. No. 6.5;
  Matches   97;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        231 CAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGG 290
              | |  ||| ||   |||| ||     ||      |  | |||  | ||  | |   | |||
Db        101 CCAGGGGGGAACCTTTGAACAAGAGGGTCTCCAGCATCTTTGAAATGCTCAGTTGATGGG 160

Qy        291 CAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCA 350
              |   | |||  ||     |||| || |||    |  | |   ||  |      ||  |
Db        161 GAAAACAAACTTTGTAGTCATTTGGCTAGGGAAATTGGCTTTAACCCCTCCTCCTACTCT 220

Qy        351 GAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGT 410
                | |  ||| | |  ||  |||   |  ||   |||| |||||  |||   |||
Db        221 TTGTTCCTATCCTAAAAGACTAACCATTTACCCCTACTCTACTGAAAAAGGGGCTGGTGT 280

Qy        411 AGTGCAGAGACAGTCGCA 428
              || ||   ||||||| ||
Db        281 AGGGCCACGACAGTCTCA 298


RESULT 11
AZ169114
LOCUS       AZ169114                597 bp    DNA     linear   GSS 29-AUG-2000
DEFINITION  SP_0113_A2_A09_T7A Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library Strongylocentrotus purpuratus
            genomic clone Plate=113 Col=18 Row=A, genomic survey sequence.
ACCESSION   AZ169114
VERSION     AZ169114.1  GI:8339482
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
            Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H.
            and Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
   PUBMED   10920195
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 113  row: A  column: 18
            Seq primer: T7
            Class: BAC ends
```

```
          High quality sequence stop: 597.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7668"
                     /clone="Plate=113 Col=18 Row=A"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 28;  Length 597;
  Best Local Similarity   51.2%;  Pred. No. 7.5;
  Matches   85;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        129 TAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTGCTCAGTTACGGCAGGG 188
              || |||     || |  ||||||| | |||    || | || || |     |  | |||
Db        233 TAGTCATCGTCCCTTTCTTGGTCACGATGGCCATAGTCATTGTCCCTTTCTTGGTCAGCC 292

Qy        189 AGGCTCAAAACTTTTGGCGGTTGTTGCGCAAGAAGGTAGTAGCAACCGGGCAAAGATTGA 248
              | ||  |  | |     |  || |||  ||     ||| |||| || || ||      ||
Db        293 TGTCTGTAGTCATCGTCCCTTTCTTGGTCAGCCTGGTCGTAGTCACTGGCCATTTCCTGC 352

Qy        249 CCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGT 294
              ||  | |     |  | | | | | | |  || |||| | ||| ||
Db        353 TCATCCTGTCTGTAGTCATCGTCCCTTTCCTGGTCAGCCTGGCCGT 398


RESULT 12
BB449135
LOCUS       BB449135                657 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  BB449135 RIKEN full-length enriched, 9 days embryo Mus musculus
            cDNA clone D030070B09 3' similar to D38046 Mouse mRNA for typeII
            DNA topoisomerase beta isoform, mRNA sequence.
ACCESSION   BB449135
VERSION     BB449135.2  GI:16425330
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 21, 2000 this sequence version replaced gi:9310170.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Pukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="D030070B09"
                     /dev_stage="9 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 9 days embryo"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
ORIGIN

  Query Match             8.0%;  Score 36.4;  DB 10;  Length 657;
  Best Local Similarity   50.6%;  Pred. No. 7.7;
  Matches   88;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        226 AGTAGCAACCGGGCAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAG 285
              || ||  || ||| |||  |     || | | |||  |  |   | || | |   | ||
Db        478 AGAAGATACAGGGGAAAATTACTATAGGCTGCAGAAGAGGAAGATTCACAAAACCAACAT 537

Qy        286 GCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATT 345
              |  | |||| ||   ||| |  | | | |   |||  || ||  |||| |    | | |
Db        538 GATGACAGTTCCTCTGATTCAGGAACTCCTTCAGGCCCTGATTTTAATTATATTTTAAAT 597

Qy        346 ATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAA 399
              ||      | | |||  || ||||| |||  ||  | |       ||  || | |
Db        598 ATGTCTCTATGGTCTCTTACTAAAGAAAAAGTTGAAGAACTCATTAAACAGAGA 651


RESULT 13
BG354883
LOCUS       BG354883                769 bp    mRNA    linear   EST 05-MAR-2001
DEFINITION  MBTsMLA133T7SEQ Trichinella spiralis muscle stage larvae (BC)
            Trichinella spiralis cDNA clone MBTsMLA133 5', mRNA sequence.
ACCESSION   BG354883
VERSION     BG354883.1  GI:13199069
KEYWORDS    EST.
SOURCE      Trichinella spiralis
  ORGANISM  Trichinella spiralis
            Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida;
```

Trichinellidae; Trichinella.
REFERENCE 1 (bases 1 to 769)
AUTHORS Daub,J., Connolly,B., Garate,T. and Blaxter,M.L.
TITLE A survey of genes expressed in the muscle stage larvae of the parasitic nematode Trichinella spiralis
JOURNAL Unpublished (2001)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
PCR PRimers
FORWARD: T7SEQ (GTAATACGACTCACTATAGGG)
BACKWARD: M13 Forward (CGCCAGGGTTTTCCCAGTCACGAC)
Seq primer: T7SEQ (GTAATACGACTCACTATAGGG).

FEATURES Location/Qualifiers
source 1. .769
/organism="Trichinella spiralis"
/mol_type="mRNA"
/db_xref="taxon:6334"
/clone="MBTsMLA133"
/sex="mixed"
/dev_stage="muscle stage larvae"
/clone_lib="Trichinella spiralis muscle stage larvae (BC)"
/note="Vector: Lambda ZipLox; Site_1: SalI (5'end); Site_2: NotI (3'end); The infective L1 larva of Trichinella spiralis is a nematode parasite of mammalian skeletal muscle. The library was constructed using muscle larvae of the T. spiralis isolate ISS3, and was provided by Dr Bernadette Connolly, University of Aberdeen."

ORIGIN

Query Match 8.0%; Score 36.4; DB 12; Length 769;
Best Local Similarity 50.6%; Pred. No. 8.1;
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
Qy      239 CAAAGATTGACCAGACAGGAGATTATAACCTTGCATATATTGATCAGGCGGGCAGTGCCA 298
            ||   ||  | | || |  ||| |||  |  || || |     |    |    | | | |
Db       78 CAGTTATGAATCGGATAACAGAAAATATACGCGCTTACAAATTTGGTCCATCGACTTCAA 137

Qy      299 ACGATGCCAGTATTTCGCAAGGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGGTT 358
            |||||| || |||    ||| || ||    ||  |||        ||         | ||
Db      138 TCGATGCAAGCATTGATGAAGATGATTTACCTATGACTTATGGATTTGAAGCTTTGGTTT 197

Qy      359 CTGGTAATAAAGCAAATATTACACAGTATGGTACTCAAAAAACGGCAATTGTAG 412
            ||| | ||||||  ||||||    |  |     |  ||||| | | ||||| ||
Db      198 CTGATGATAAAGGCAATATTGTTAATGAACTGATGAAAAAATCAGTAATTGGAG 251
```

RESULT 14
AZ811111
LOCUS AZ811111 491 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0077A11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0077A11 F, genomic survey sequence.
ACCESSION AZ811111
VERSION AZ811111.1 GI:12979042
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 491)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0077 row: A column: 11
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 491.

FEATURES Location/Qualifiers
source 1. .491
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0077A11"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 7.9%; Score 36.2; DB 28; Length 491;
Best Local Similarity 52.3%; Pred. No. 8.2;
Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy      270 TGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGG 329
            ||   |||   ||     | || ||| || | | | |||||| |||     || |||  |
Db      110 TGAGGATAAATATTTTCAGAGCTGTGTCAGCAAAGGCAGTATATCGTTTCATGTTTAGAG 169

Qy      330 TAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGG 389
            ||    ||||||  || |  | | | |   |||||   ||| ||  |  | | | |
Db      170 TACATGTGCGATAGTTCTGGAAATATGGATTGGTATACAAGAAAGAAAAAGAAAAAAGAA 229

Qy      390 TACTCAAAAAACGGCAATTGTAGTGCAGAGACA 422
              |     ||||| |  ||  | |  | | ||| |
Db      230 AAAAAGAAAAAAGAGAAAAGAATAGAAAAGAAA 262
```

RESULT 15
AZ811109
LOCUS AZ811109 532 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0077A09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0077A09 F, genomic survey sequence.
ACCESSION AZ811109
VERSION AZ811109.1 GI:12979038
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 532)

```
  AUTHORS   Dunn,D., Aoyagi;A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0077  row: A  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 532.
FEATURES            Location/Qualifiers
     source         1. .532
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UUGC2M0077A09"
                    /sex="Male"
                    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                    /clone_lib="Mouse 10kb plasmid UUGC1M library"
                    /note="Vector: PWD42nv; Purified genomic DNA from M.
                    musculus C57BL/6J (male) was obtained  from the Jackson
                    Laboratory Mouse DNA Resource
                    (http://www.jax.org/resources/documents/dnares/). The DNA
                    was hydrodynamically sheared by repeated passage through a
                    0.005 inch orifice at constant velocity. The sheared DNA
                    was blunt end-repaired with T4 DNA polymerase and T4
                    polynucleotide kinase. Adaptor oligonucleotides were
                    ligated to the blunt ends in high  molar excess. The
                    adaptored DNA was purified and size-selected for a 9.5 to
                    10.5 kb range using preparative agarose gel
                    electrophoresis. Vector DNA was prepared from a derivative
                    of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                    inducible derivative of plasmid R1. The vector was ligated
                    with adaptors complementary to the insert adaptors and
                    purified. The sheared, adaptored mouse DNA was annealed to
                    adaptored vector DNA, and transformed into
                    chemically-competent E. coli XL10-Gold (Stratagene) cells
                    and selected for ampicillin resistance."
ORIGIN

  Query Match             7.9%;  Score 36.2;  DB 28;  Length 532;
  Best Local Similarity   52.3%;  Pred. No. 8.4;
  Matches   80;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        270 TGCATATATTGATCAGGCGGGCAGTGCCAACGATGCCAGTATTTCGCAAGGTGCTTATGG 329
              ||   |||   ||     | || ||| || | | | |||||| |||      || |||  |
Db         92 TGAGGATAAATATTTTCAGAGCTGTGTCAGCAAAGGCAGTATATCGTTTCATGTTTAGAG 151

Qy        330 TAATACTGCGATGATTATCCAGAAAGGTTCTGGTAATAAAGCAAATATTACACAGTATGG 389
              ||    |||||| || |  | | | |   |||||   ||| ||  | | | | |
Db        152 TACATGTGCGATAGTTCTGGAAATATGGATTGGTATACAAGAAAGAAAAAGAAAAAAGAA 211

Qy        390 TACTCAAAAAACGGCAATTGTAGTGCAGAGACA 422
                |    ||||| |  ||  | |  | | ||| |
Db        212 AAAAAGAAAAAAGAGAAAAGAATAGAAAAGAAA 244


Search completed: March 18, 2004, 20:11:20
Job time : 2038.55 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:10:42 ; Search time 243.087 Seconds
(without alignments)
8558.533 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		29.4	61.2	1797	3	LEIMSP52A	L19563 Leishmania
2		29.4	61.2	1860	3	LIGP63GEN	Z83677 L.infantum
3		29.4	61.2	2104	3	LDO495002	AJ495002 Leishmani
4		29.4	61.2	2106	3	LDO495006	AJ495006 Leishmani
5		29.4	61.2	2106	3	LDO495007	AJ495007 Leishmani
6		29.4	61.2	2107	3	LDO495005	AJ495005 Leishmani
7		29.4	61.2	2109	3	LDO495003	AJ495003 Leishmani
8		29.4	61.2	2109	3	LDO495004	AJ495004 Leishmani
9		29.4	61.2	2892	3	LEIGP63D	M80672 Leishmania
10		29.4	61.2	2966	3	LEIGP63B	M80669 Leishhmania
11		29.4	61.2	3047	3	LEIGP63A	M28527 L.chagasi m
12		29.4	61.2	3047	3	LIU48798	U48798 Leishmania
13		29.4	61.2	3105	3	LEIGPAA	M60048 L.donovani.
14		28.4	59.2	2621	3	LEIGP63C	M80671 Leishmania
15		27.8	57.9	2050	3	LTU495010	AJ495010 Leishmani
16		27.8	57.9	2058	3	LAE532831	AJ532831 Leishmani
17		27.8	57.9	2077	3	LTR495009	AJ495009 Leishmani
18		27.8	57.9	2090	3	LAR532830	AJ532830 Leishmani
19		27.8	57.9	2125	3	LTR495008	AJ495008 Leishmani
20		27.8	57.9	2161	3	LMGP63	Y00647 Leishmania
21		27.8	57.9	2187	3	LMGP63C1	X64394 L.mexicana
22		27.8	57.9	3229	3	AF039721	AF039721 Leishmani
23	c	27.8	57.9	11523	1	AE014950	AE014950 Streptoco
24	c	26.8	55.8	57112	2	AC019976	AC019976 Drosophil
25	c	26.8	55.8	88031	2	AC004378	AC004378 Drosophil
26		26.8	55.8	124700	3	AC005558	AC005558 Drosophil
27		26.8	55.8	126807	9	HS391O22	AL031577 Human DNA
28	c	26.8	55.8	171774	3	AC009909	AC009909 Drosophil
29	c	26.8	55.8	295134	3	AE003582	AE003582 Drosophil
30		26.4	55.0	1800	3	LEIMSPS4A	L19562 Leishmania
31		26.2	54.6	2046	3	LEIGP63E	L46798 Leishmania
32	c	25.8	53.8	238365	2	AC107269	AC107269 Rattus no
33		24.6	51.3	1964	3	LIGP63	Y08156 L.infantum
34		24.6	51.3	2544	6	AX568314	AX568314 Sequence
35	c	24.6	51.3	3636	6	AR218846	AR218846 Sequence
36	c	24.6	51.3	3636	6	BD003758	BD003758 Polynucle
37		24.6	51.3	11495	1	AE007386	AE007386 Streptoco
38		24.6	51.3	12092	1	AE008447	AE008447 Streptoco
39		24.6	51.3	147006	2	SPNEU1904	AL449926 Streptoco
40		24.6	51.3	349980	6	AX571761	AX571761 Sequence
41		24.6	51.3	349980	6	AX571762	AX571762 Sequence
42	c	24.4	50.8	2240	8	AY118173	AY118173 Ustilago
43		24.2	50.4	144239	10	AL627347	AL627347 Mouse DNA
44	c	24.2	50.4	182340	2	AC099639	AC099639 Mus muscu
45	c	24.2	50.4	223974	2	AC106343	AC106343 Rattus no

ALIGNMENTS

RESULT 1
LEIMSP52A
LOCUS LEIMSP52A 1797 bp DNA linear INV 28-NOV-1994
DEFINITION Leishmania donovani glycoprotein 63 (mspS2) gene, complete cds.
ACCESSION L19563
VERSION L19563.1 GI:308887
KEYWORDS glycoprotein 63; surface protein; virulence factor; zinc protease.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 1797)
AUTHORS Roberts,S.C., Swihart,K.G., Agey,M.W., Ramamoorthy,R., Wilson,M.E. and Donelson,J.E.
TITLE Sequence diversity and organization of the msp gene family encoding

gp63 of Leishmania chagasi
JOURNAL Mol. Biochem. Parasitol. 62 (2), 157-171 (1993)
MEDLINE 94187791
PUBMED 8139613
COMMENT Original source text: Leishmania donovani (sub_species chagasi) stationary stage promastigote DNA.

FEATURES Location/Qualifiers
source 1. .1797
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
/dev_stage="stationary stage promastigote"
gene 1. .1797
/gene="mspS2"
CDS 1. .1797
/gene="mspS2"
/note="propeptide"
/codon_start=1
/product="glycoprotein 63"
/protein_id="AAA53688.1"
/db_xref="GI:308888"
/translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEG
DVLAWAATCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSDTFFTD
KRMLHNVGKIRGKPHNAPVINSSTAVAKAREQYGCDTLEYLEMEDEGGAGSAGSHIKM
RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
DGITKWPAMFCNSDDALRCPTSRLSLGACSLATYQSLPPYWQYFTDPSLAGISAFMDY
CPVVVPFGNGSCAQNASKVMAAVQAFNVFSDAARCIDGAFRPKTTETVTNSYAGLCAN
VLCDTATRTYSVQVRGGSGYASCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNVQ
AAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
sig_peptide 1. .39
/gene="mspS2"
mat_peptide 97. .1794
/gene="mspS2"
/product="glycoprotein 63"

ORIGIN

Query Match 61.2%; Score 29.4; DB 3; Length 1797;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        751 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 797
```

RESULT 2
LIGP63GEN
LOCUS LIGP63GEN 1860 bp DNA linear INV 16-DEC-1996
DEFINITION L.infantum GP63 gene.
ACCESSION Z83677
VERSION Z83677.1 GI:1743284
KEYWORDS gp63 gene; protease.
SOURCE Leishmania infantum
ORGANISM Leishmania infantum
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
REFERENCE 1 (bases 1 to 1860)
AUTHORS Ortiz,G., Martinez,P. and Segovia,M.
TITLE Characterization of GP63 of Leishmania infantum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1860)
AUTHORS Ortiz,G.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1996) Ortiz G., Facultad de Medicina. Universidad de Murcia, Genetica y Microbiologia, Campus de Espinardo, Murcia, Murcia, SPAIN, 30100

FEATURES Location/Qualifiers
source 1. .1860
/organism="Leishmania infantum"
/mol_type="genomic DNA"
/strain="cutaneous"
/db_xref="taxon:5671"
/dev_stage="promastigote"
CDS 16. .1815
/function="protease"
/codon_start=1
/product="GP63"
/protein_id="CAB06018.1"
/db_xref="GI:1743285"
/translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGIIKSYAGLCA
NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"

ORIGIN

Query Match 61.2%; Score 29.4; DB 3; Length 1860;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        766 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 812
```

RESULT 3
LDO495002
LOCUS LDO495002 2104 bp DNA linear INV 22-JUL-2002
DEFINITION Leishmania donovani mspC gene for GP63, strain MHOM/KE/0000/Neal-R1.
ACCESSION AJ495002
VERSION AJ495002.1 GI:21954455
KEYWORDS GP63; major surface protease; mspC gene.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
REFERENCE 1
AUTHORS Mauricio,I.L., Stothard,J.R. and Miles,M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2104)
AUTHORS Mauricio,I.L.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM

FEATURES Location/Qualifiers
source 1. .2104
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="MHOM/KE/0000/Neal-R1"
/db_xref="taxon:5661"
/country="Kenya"
gene 15. .1940
/gene="mspC"
CDS 15. .1940
/gene="mspC"
/note="constitutive major surface protease"
/codon_start=1
/allele="mspCLdA1"
/product="GP63"
/protein_id="CAD42811.1"
/db_xref="GI:21954456"
/db_xref="GOA:Q8MNZ5"
/db_xref="SPTREMBL:Q8MNZ5"

```
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVACHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDSAYHCARVGQRVNNHAGAIATCTAEDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWNVTDMVDEICGDFKVPQAHITEGFSNTDFVMYVASVPSEK
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILLVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVRGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQAN
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     15. .131
                     /gene="mspC"
     mat_peptide     132. .1823
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 2104;
  Best Local Similarity   76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        765 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 811


RESULT 4
LDO495006
LOCUS       LDO495006               2106 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain MCAN/IQ/1977/Bumm3.
ACCESSION   AJ495006
VERSION     AJ495006.1  GI:21954463
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2106)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MCAN/IQ/1977/Bumm3"
                     /db_xref="taxon:5661"
                     /country="Iraq"
     gene            14. .1939
                     /gene="mspC"
     CDS             14. .1939
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdG2"
                     /product="GP63"
                     /protein_id="CAD42815.1"
                     /db_xref="GI:21954464"
                     /db_xref="GOA:Q8MM48"
                     /db_xref="SPTREMBL:Q8MM48"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMHPRVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTADDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWKVTDMVDEICGDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILVVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCGTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVCGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQAN
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     14. .130
                     /gene="mspC"
     mat_peptide     131. .1822
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 2106;
  Best Local Similarity   76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        764 TACGACCAGCTGGTGACACGTGTCGTCACGCACGAGATGGCGCACGC 810


RESULT 5
LDO495007
LOCUS       LDO495007               2106 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain
            MHOM/SU/1984/Marz-Krim.
ACCESSION   AJ495007
VERSION     AJ495007.1  GI:21954465
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2106)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MHOM/SU/1984/Marz-Krim"
                     /db_xref="taxon:5661"
     gene            14. .1939
                     /gene="mspC"
     CDS             14. .1939
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdH"
                     /product="GP63"
                     /protein_id="CAD42816.1"
                     /db_xref="GI:21954466"
                     /db_xref="GOA:Q8MM48"
                     /db_xref="SPTREMBL:Q8MM48"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMHPRVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTADDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWKVTDMVDEICGDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILVVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCGTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVCGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQAN
```

```
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     14. .130
                     /gene="mspC"
     mat_peptide     131. .1822
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 2106;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        764 TACGACCAGCTGGTGACACGTGTCGTCACGCACGAGATGGCGCACGC 810


RESULT 6
LDO495005
LOCUS       LDO495005               2107 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain
            MCAN/IQ/1981/Sukkar2.
ACCESSION   AJ495005
VERSION     AJ495005.1  GI:21954461
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2107)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2107
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MCAN/IQ/1981/Sukkar2"
                     /db_xref="taxon:5661"
                     /country="Iraq"
     gene            14. .1939
                     /gene="mspC"
     CDS             14. .1939
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdG1"
                     /product="GP63"
                     /protein_id="CAD42814.1"
                     /db_xref="GI:21954462"
                     /db_xref="GOA:Q8MNZ2"
                     /db_xref="SPTREMBL:Q8MNZ2"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMHPRVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTADDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWKVTDMVDEICGDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILVVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCGTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVCGSMDYVNCTPGLRVELSXVSSAFEEGGYITCPPYVEVCQAN
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     14. .130
                     /gene="mspC"
     mat_peptide     131. .1822
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 2107;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        764 TACGACCAGCTGGTGACACGTGTCGTCACGCACGAGATGGCGCACGC 810


RESULT 7
LDO495003
LOCUS       LDO495003               2109 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain
            MHOM/KE/1954/LRC-L53.
ACCESSION   AJ495003
VERSION     AJ495003.1  GI:21954457
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2109)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2109
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MHOM/KE/1954/LRC-L53"
                     /db_xref="taxon:5661"
                     /country="Kenya"
     gene            15. .1940
                     /gene="mspC"
     CDS             15. .1940
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdA2"
                     /product="GP63"
                     /protein_id="CAD42812.1"
                     /db_xref="GI:21954458"
                     /db_xref="GOA:Q8MNZ4"
                     /db_xref="SPTREMBL:Q8MNZ4"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVACHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTAEDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWNVTDMVDEICGDFKVPQAHITEGFSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILLVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCME
                     QNITKWPAMFCNVSVDVVRCPTSRLMLGTCGIRGYSTPFSPYWQYFTNISLGGYSPFL
                     DYCPFVIGYGDGSCNQDASLATGFFGAFNVFSDAARCIDGAFRPKNRTAADGYYAGLC
                     ANVRCDTATRTYSVQVRGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQAN
                     VKGAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     15. .131
                     /gene="mspC"
     mat_peptide     132. .1823
                     /gene="mspC"
                     /product="GP63"
ORIGIN
```

```
  Query Match            61.2%;  Score 29.4;  DB 3;  Length 2109;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        765 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 811


RESULT 8
LDO495004
LOCUS       LDO495004               2109 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania donovani mspC gene for GP63, strain
            MHOM/CN/0000/WangJie1.
ACCESSION   AJ495004
VERSION     AJ495004.1  GI:21954459
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2109)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2109
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /strain="MHOM/CN/0000/WangJie1"
                     /db_xref="taxon:5661"
                     /country="China"
     gene            15. .1940
                     /gene="mspC"
     CDS             15. .1940
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLdF"
                     /product="GP63"
                     /protein_id="CAD42813.1"
                     /db_xref="GI:21954460"
                     /db_xref="GOA:Q8MNZ3"
                     /db_xref="SPTREMBL:Q8MNZ3"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTAEDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWNVTDMVDEICGDFKVPQAHITEGFSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSGTFFTE
                     ILLVTQMMNIRGKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNESEVTMRCPTSRLMVGTCGIRGYSTPFSLYWQYFTNASLGGYSPFL
                     DYCPFVIGYSDGSCNQDASLAAGFFSAFNVFSDAARCIDGAFRPKNRTAANGYYAGLC
                     ANVRCDTATRTYSVQVRGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGN
                     VQAAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWL
                     LLLTCPWWCCKFGGLPT"
     sig_peptide     15. .131
                     /gene="mspC"
     mat_peptide     132. .1823
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match            61.2%;  Score 29.4;  DB 3;  Length 2109;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        765 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 811


RESULT 9
LEIGP63D
LOCUS       LEIGP63D                2892 bp    mRNA    linear   INV 26-APR-1993
DEFINITION  Leishmania donovani chagasi surface protease (log glycoprotein 63,
            GP63) mRNA, complete cds.
ACCESSION   M80672
VERSION     M80672.1  GI:159328
KEYWORDS    log glycoprotein 63; surface protease.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 2892)
  AUTHORS   Ramamoorthy,R., Donelson,J.E., Paetz,K.E., Maybodi,M., Roberts,S.C.
            and Wilson,M.E.
  TITLE     Three distinct RNAs for the surface protease gp63 are
            differentially expressed during development of Leishmania donovani
            chagasi promastigotes to an infectious form
  JOURNAL   J. Biol. Chem. 267 (3), 1888-1895 (1992)
  MEDLINE   92112918
   PUBMED   1370484
COMMENT     Original source text: Leishmania donovani (sub_species chagasi)
            cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .2892
                     /organism="Leishmania donovani"
                     /mol_type="mRNA"
                     /sub_species="chagasi"
                     /db_xref="taxon:5661"
     gene            1. .2892
                     /gene="GP63"
     CDS             114. .1913
                     /gene="GP63"
                     /note="log glycoprotein 63"
                     /codon_start=1
                     /product="surface protease"
                     /protein_id="AAA29238.1"
                     /db_xref="GI:159329"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
                     ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
                     ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
                     CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGIIKSYAGLCA
                     NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAPEEGGYITCPPYVEVCQGNV
                     QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
     sig_peptide     114. .230
                     /gene="GP63"
     mat_peptide     405. .1910
                     /gene="GP63"
                     /product="surface protease"
                     /note="log glycoprotein 63"
     polyA_site      2892
                     /gene="GP63"
ORIGIN

  Query Match            61.2%;  Score 29.4;  DB 3;  Length 2892;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        864 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 910
```

```
RESULT 10
LEIGP63B
LOCUS       LEIGP63B                2966 bp    mRNA    linear   INV 26-APR-1993
DEFINITION  Leishhmania donovani chagasi surface protease (stationary
            glycoprotein 63, GP63) mRNA, complete cds.
ACCESSION   M80669
VERSION     M80669.1  GI:159324
KEYWORDS    stationary glycoprotein 63; surface protease.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 2966)
  AUTHORS   Ramamoorthy,R., Donelson,J.E., Paetz,K.E., Maybodi,M., Roberts,S.C.
            and Wilson,M.E.
  TITLE     Three distinct RNAs for the surface protease gp63 are
            differentially expressed during development of Leishmania donovani
            chagasi promastigotes to an infectious form
  JOURNAL   J. Biol. Chem. 267 (3), 1888-1895 (1992)
  MEDLINE   92112918
   PUBMED   1370484
COMMENT     Original source text: Leishmania donovani (sub_species chagasi)
            cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .2966
                     /organism="Leishmania donovani"
                     /mol_type="mRNA"
                     /sub_species="chagasi"
                     /db_xref="taxon:5661"
     gene            1. .2966
                     /gene="GP63"
     CDS             129. .1928
                     /gene="GP63"
                     /note="stationary glycoprotein 63"
                     /codon_start=1
                     /product="surface protease"
                     /protein_id="AAA29236.1"
                     /db_xref="GI:159325"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
                     ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWAATCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
                     ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
                     CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGQIKSYAGLCA
                     NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
                     QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
     sig_peptide     129. .245
                     /gene="GP63"
     mat_peptide     420. .1925
                     /gene="GP63"
                     /product="surface protease"
                     /note="stationary glycoprotein 63"
     polyA_site      2966
                     /gene="GP63"
ORIGIN

  Query Match             61.2%;  Score 29.4;  DB 3;  Length 2966;
  Best Local Similarity   76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        879 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 925


RESULT 11
LEIGP63A
LOCUS       LEIGP63A                3047 bp    DNA     linear   INV 26-APR-1993
DEFINITION  L.chagasi major surface glycoprotein (gp63) gene, complete cds.
ACCESSION   M28527
VERSION     M28527.1  GI:159322
KEYWORDS    glycoprotein; protease; surface antigen.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 3047)
  AUTHORS   Miller,R.A., Reed,S.G. and Parsons,M.
  TITLE     Leishmania gp63 molecule implicated in cellular adhesion lacks an
            Arg-Gly-Asp sequence
  JOURNAL   Mol. Biochem. Parasitol. 39 (2), 267-274 (1990)
  MEDLINE   90205976
   PUBMED   2320059
COMMENT     Original source text: L.chagasi (isolate MHOM/BR/82/BA-2C1a) DNA,
            clones pLc63-[1 and 2].
            Draft entry and computer-readable sequence for [1] kindly provided
            by M.Parsons, 03-OCT-1989.
FEATURES             Location/Qualifiers
     source          1. .3047
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /sub_species="chagasi"
                     /db_xref="taxon:5661"
     CDS             496. .2295
                     /note="gp63 protein"
                     /codon_start=1
                     /protein_id="AAA29235.1"
                     /db_xref="GI:159323"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
                     ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
                     ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
                     CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGIIKSYAGLCA
                     NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
                     QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
ORIGIN

  Query Match             61.2%;  Score 29.4;  DB 3;  Length 3047;
  Best Local Similarity   76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db       1246 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 1292


RESULT 12
LIU48798
LOCUS       LIU48798                3047 bp    DNA     linear   INV 15-JAN-1998
DEFINITION  Leishmania infantum surface glycoprotein gp63 (gp63) gene, complete
            cds.
ACCESSION   U48798
VERSION     U48798.1  GI:1213329
KEYWORDS    .
SOURCE      Leishmania infantum
  ORGANISM  Leishmania infantum
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 3047)
  AUTHORS   Gonzalez-Aseguinolaza,G., Almazan,F., Rodriguez,J.F., Marquet,A.
            and Larraga,V.
  TITLE     Cloning of the gp63 surface protease of Leishmania infantum.
            Differential post-translational modifications correlated with
            different infective forms
  JOURNAL   Biochim. Biophys. Acta 1361 (1), 92-102 (1997)
  MEDLINE   97388259
   PUBMED   9247093
REFERENCE   2  (bases 1 to 3047)
```

```
  AUTHORS   Gonzalez-Aseguinolaza,G., Almazan,F., Rodriguez-Aguirre,J.F.,
            Marquet,A. and Larraga,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-1996) Vicente Larraga, Proteins Structure, Center
            of Biological Research, Spanish Research Council, Velazquez 144,
            Madrid 28006, Spain
FEATURES             Location/Qualifiers
     source          1. .3047
                     /organism="Leishmania infantum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5671"
     gene            496. .2295
                     /gene="gp63"
     CDS             496. .2295
                     /gene="gp63"
                     /note="protease"
                     /codon_start=1
                     /product="surface glycoprotein gp63"
                     /protein_id="AAB96339.1"
                     /db_xref="GI:1213330"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQHIKRRLGGVDICTAEDILTDEKRDILVKHLIPQ
                     ALQLHTERLKVRQVQDKWKVTGMGDDVCSDFKVPPAHITDGLSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVGFFEG
                     ARILESISNVRHKDFDVPVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKM
                     RNAQDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCME
                     RNITKWPAMFCNENEVTMRCPTSRLSLGKCGVTRHPDLPPYWQYFTDPSLAGISAFMD
                     CCPVVEPYGDGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTSHGIIKSYAGLCA
                     NVRCDTATRTYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNV
                     QAAKDGGNAAAGRRGPRAAATALLVAALLAVAL"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 3047;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db       1246 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 1292


RESULT 13
LEIGPAA
LOCUS       LEIGPAA                 3105 bp    DNA     linear   INV 26-APR-1993
DEFINITION  L.donovani.
ACCESSION   M60048
VERSION     M60048.1  GI:159334
KEYWORDS    glycoprotein 63.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 3105)
  AUTHORS   Webb,J.R., Button,L.L. and McMaster,W.R.
  TITLE     Heterogeneity of the genes encoding the major surface glycoprotein
            of Leishmania donovani
  JOURNAL   Mol. Biochem. Parasitol. 48 (2), 173-184 (1991)
  MEDLINE   92107220
   PUBMED   1762629
COMMENT     Original source text: L.donovani DNA.
FEATURES             Location/Qualifiers
     source          1. .3105
                     /organism="Leishmania donovani"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5661"
     gene            101. .1873
                     /gene="gp63"
     CDS             101. .1873
                     /gene="gp63"
                     /codon_start=1
                     /product="glycoprotein 63"
                     /protein_id="AAA29244.1"
                     /db_xref="GI:159335"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLSYVTLGAAPTVVRAANWGALRIAVST
                     EDLTDSAYHCARVGQRISTRDGRFAICTAEDILTDEKRDILVKYLIPQALQLHTERLK
                     VRQVQDKWKVTGMGNEICGHFKVPPAHITDGLSNTDFVMYVASVPSEGDVLAWATTCQ
                     VFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHALGFSVVFFRDARILESISNV
                     RHKDFDVPVINSSTAVAKAREQYGCGTLEYLEMEDQGGAGSAGSHIKMRNAQDELMAP
                     ASDAGYYSALTMAIFQDLGFYQADFSKAEEMPWGRNAGCAFLSEKCMEDGITKWPAMF
                     CNENEVTMRCHTGRLSLGVCGLSSSDIPLPPYWQYFTDPLLAGISAFMDYCPVVVPFG
                     DGSCAQRASEAGAPFKGFNVFSDAARCIDGAFRPKTTETVTNSYAGLCANVRCDTATR
                     TYSVQVHGGSGYANCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQGNVQAAKDGGNA
                     AAGRRGPRAAATALLVAALLAVAL"
ORIGIN

  Query Match           61.2%;  Score 29.4;  DB 3;  Length 3105;
  Best Local Similarity  76.6%;  Pred. No. 2.8;
  Matches   36;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || |||||||| || ||||| || || || || ||||| || ||
Db        821 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACGC 867


RESULT 14
LEIGP63C
LOCUS       LEIGP63C                2621 bp    mRNA    linear   INV 26-APR-1993
DEFINITION  Leishmania donovani chagasi surface protease (constitutive
            glycoprotein 63, GP63) mRNA, complete cds.
ACCESSION   M80671
VERSION     M80671.1  GI:159326
KEYWORDS    constitutive glycoprotein 63; surface protease.
SOURCE      Leishmania donovani
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 2621)
  AUTHORS   Ramamoorthy,R., Donelson,J.E., Paetz,K.E., Maybodi,M., Roberts,S.C.
            and Wilson,M.E.
  TITLE     Three distinct RNAs for the surface protease gp63 are
            differentially expressed during development of Leishmania donovani
            chagasi promastigotes to an infectious form
  JOURNAL   J. Biol. Chem. 267 (3), 1888-1895 (1992)
  MEDLINE   92112918
   PUBMED   1370484
COMMENT     Original source text: Leishmania donovani (sub_species chagasi)
            cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .2621
                     /organism="Leishmania donovani"
                     /mol_type="mRNA"
                     /sub_species="chagasi"
                     /db_xref="taxon:5661"
     gene            1. .2621
                     /gene="GP63"
     CDS             132. .2051
                     /gene="GP63"
                     /note="constitutive glycoprotein 63"
                     /codon_start=1
                     /product="surface protease"
                     /protein_id="AAA29237.1"
                     /db_xref="GI:159327"
                     /translation="MSVDSSSTHRHRSVAARLVRLAAAGAAVIAAVGTAAAWAHAGAV
                     QHRCIHDAMQARVRQSVARHHTAPGAVSAVGLPYVTLDTAAAADRRPGSAPTVVRAAN
                     WGALRIAVSTEDLTDPAYHCARVGQRVNNHAGAIATCTAEDILTDEKRDILVKYLIPQ
                     ALQLHTERLKVRQVQDKWNVTGMVDEICGDFKVPPAHITEGFSNTDFVMYVASVPSEE
                     GVLAWATTCQVFSDGHPAVGVINIPAANIASRYDQLVTRVVTHEMAHVGFSGTFFTEI
                     LLVTQMMNIRGKDFNVSVINSSTAVAKAREQYGCDTLEYLEIEDQGGAGSAGSHIKMR
                     NAKDELMAPAAAAGYYSALTMAIFQDLGFYQADFSKAEVMPWGRNAGCAFLSEKCMER
                     NITKWPAMFCNENEVTMRCPTSRLMVGTCGIRGYSTPFSLYWQYFTNASLGLLAFLDY
                     CPFVIGYSDGSCNQDASLAAGFFSAFNVFSDAARCIDGAFRPKNRTAANGYYAGLCAN
                     VRCDTATRTYSVQVRGSMDYVNCTPGLRVELSTVSSAFEEGGYITCPPYVEVCQANVK
                     GAKDFAGDSDSSSSAGDAADRAAMQRWNDRMAGLATAAMVLLGMVLSLMALVVVWLLL
                     LTCPWWCCKFGGLPT"
```

```
     sig_peptide     132. .248
                     /gene="GP63"
     mat_peptide     423. .2048
                     /gene="GP63"
                     /product="surface protease"
                     /note="constitutive glycoprotein 63"
     polyA_site      2621
                     /gene="GP63"
ORIGIN

  Query Match             59.2%;  Score 28.4;  DB 3;  Length 2621;
  Best Local Similarity   76.1%;  Pred. No. 6.8;
  Matches   35;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              || || |||||||| || ||||| || || || || ||||| || |
Db        882 TACGACCAGCTGGTGACGCGTGTCGTCACGCACGAGATGGCGCACG 927


RESULT 15
LTU495010
LOCUS       LTU495010               2050 bp    DNA     linear   INV 22-JUL-2002
DEFINITION  Leishmania turanica mspC gene for GP63, strain
            MRHO/SU/1983/MARZ-051.
ACCESSION   AJ495010
VERSION     AJ495010.1  GI:21954471
KEYWORDS    GP63; major surface protease; mspC gene.
SOURCE      Leishmania turanica
  ORGANISM  Leishmania turanica
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1
  AUTHORS   Mauricio,I.L., Stothard,J.R. and Miles,M.A.
  TITLE     Genetic diversity in the Leishmania donovani complex
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2050)
  AUTHORS   Mauricio,I.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
            Diseases, London School of Hygiene and Trop. Med., Keppel Street,
            WC1E 7HT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2050
                     /organism="Leishmania turanica"
                     /mol_type="genomic DNA"
                     /strain="MRHO/SU/1983/MARZ-051"
                     /db_xref="taxon:62297"
     gene            15. .1949
                     /gene="mspC"
     CDS             15. .1949
                     /gene="mspC"
                     /note="constitutive major surface protease"
                     /codon_start=1
                     /allele="mspCLtuA1"
                     /product="GP63"
                     /protein_id="CAD42819.1"
                     /db_xref="GI:21954472"
                     /db_xref="GOA:Q8MNY9"
                     /db_xref="SPTREMBL:Q8MNY9"
                     /translation="MSVDSSSTHRRRCVAARLVRLAAAGVAVTVAVGTAAAWAHAGAL
                     QHRCVHDAMQARVRQSVADHHTAPGAVSAVGLPYVTLDAAHTAAAADPRPGSARGVER
                     DVNWGALRIAVSIEDLTDPXYHCVRVGQXVNNHAGAIVTCTAEDILTDEKRDILVKHL
                     IPRALXLHTERLKVQQVQGTWKVTGMVDEICGDFKVPQAHITKGFSNTDFVMYVASVP
                     SEEGVLAWATTCQVFSDGHPAVGVINIPAANIVSRYDQLVTRVVTHEMAHALGFSGPF
                     FEDARIVANISNVRGKNFDVPVINSSTAVAKAREQYGCDTLEYLEVEDQGGAGSAGSH
                     IKMRNAQDELMAPAAAAGYYTALTMAIFQDLGFYQADFSKAEVMLWGRNAGCAFLTDK
                     CMEQNITQWPAMFCNESEDAIRCPTSRLILGTCGIREYETPLPLYWQYFTNASLGGYS
                     PFLDYCPSVIGYAHGXCNQDASSAAELLAAFNVFSDAARCIDGAFTPKNRTAANGYYG
                     GLCANVQCDTATRTYSVQVHGSMDYANCTPGLRVELSTVSNTFEEGGYITCPPYVEVC
                     QGNVKGAKDFDGDSDSSSSAGDDADKAAMERWNDRMAGLATAVTVLLGIVLSLMALVV
                     VWLLLVSCPRWCCKVGGLPT"
      sig_peptide    15. .131
                     /gene="mspC"
     mat_peptide     132. .1832
                     /gene="mspC"
                     /product="GP63"
ORIGIN

  Query Match             57.9%;  Score 27.8;  DB 3;  Length 2050;
  Best Local Similarity   74.5%;  Pred. No. 11;
  Matches   35;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              || || ||||| || || ||||| || || || || ||||| || ||
Db        774 TACGACCAGCTCGTCACGCGTGTCGTCACGCACGAGATGGCGCACGC 820


Search completed: March 18, 2004, 15:02:39
Job time : 246.087 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 06:07:52 ; Search time 30.3584 Seconds
(without alignments)
6716.877 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	48	100.0	48	3	AAC64621	Aac64621 Leishmani
c	2	48	100.0	78	3	AAC64609	Aac64609 AgfA (SEF
	3	48	100.0	78	3	AAC64610	Aac64610 AgfA (SEF
	4	48	100.0	456	3	AAC64628	Aac64628 AgfA::PT3
	5	48	100.0	456	3	AAC64622	Aac64622 AgfA::PT3
	6	48	100.0	456	3	AAC64626	Aac64626 AgfA::PT3
	7	48	100.0	456	3	AAC64630	Aac64630 AgfA::PT3
	8	48	100.0	456	3	AAC64629	Aac64629 AgfA::PT3
	9	48	100.0	456	3	AAC64625	Aac64625 AgfA::PT3
	10	48	100.0	456	3	AAC64631	Aac64631 AgfA::PT3
	11	48	100.0	456	3	AAC64623	Aac64623 AgfA::PT3
	12	48	100.0	456	3	AAC64627	Aac64627 AgfA::PT3
	13	48	100.0	456	3	AAC64624	Aac64624 AgfA::PT3
	14	43.2	90.0	48	3	AAC64616	Aac64616 S. enteri
	15	40.6	84.6	78	3	AAC64606	Aac64606 SefA (SEF
c	16	40.6	84.6	78	3	AAC64605	Aac64605 SefA (SEF
	17	27.8	57.9	2547	7	ACA48284	Aca48284 Prokaryot
	18	24.6	51.3	1780	2	AAV65244	Aav65244 S. pneumo
	19	24.6	51.3	2544	7	ABX06472	Abx06472 S. pneumo
	20	24.6	51.3	2547	7	ACA49840	Aca49840 Prokaryot
c	21	24.6	51.3	3636	2	AAV52211	Aav52211 Streptoco
	22	24.6	51.3	110000	7	ABS56454_06	Continuation (7 of
	23	23	47.9	420	6	ABQ69493	Abq69493 Listeria
	24	23	47.9	420	6	ABQ67505	Abq67505 Listeria
c	25	23	47.9	436	8	ACH39062	Ach39062 Human foe
c	26	23	47.9	810	5	AAH52245	Aah52245 Human AFP
	27	23	47.9	951	4	ABL21387	Abl21387 Drosophil
c	28	23	47.9	1722	6	ABQ76709	Abq76709 Human mac
	29	23	47.9	2940	2	AAT36490	Aat36490 Lactococc
	30	23	47.9	3333	4	ABL21386	Abl21386 Drosophil
c	31	23	47.9	4850	7	ABQ83879	Abq83879 Human MDD
	32	23	47.9	18949	4	AAK70445	Aak70445 Human imm
c	33	23	47.9	63020	6	ABQ67197_11	Continuation (12 o
c	34	23	47.9	110000	6	ABQ69245_12	Continuation (13 o
c	35	22.8	47.5	1323	3	AAF13368	Aaf13368 Aspergill
	36	22.6	47.1	1767	7	ACA37452	Aca37452 Prokaryot
	37	22.6	47.1	3122	2	AAQ79913	Aaq79913 Lys-amino
	38	22.4	46.7	1014	4	AAF44754	Aaf44754 Nucellin-
	39	22.4	46.7	1014	9	ADB94941	Adb94941 PCD pathw
	40	22.4	46.7	1925	9	ADB94853	Adb94853 PCD pathw
	41	22.2	46.2	912	5	AAH67547	Aah67547 C glutami
	42	22.2	46.2	1035	4	AAF71443	Aaf71443 Corynebac
	43	22.2	46.2	44147	6	ABK84481	Abk84481 Human cDN
	44	22.2	46.2	44147	9	ADD14691	Add14691 Human src
c	45	22.2	46.2	113604	8	ADA42607	Ada42607 Human LQT

ALIGNMENTS

RESULT 1
AAC64621
ID AAC64621 standard; DNA; 48 BP.
XX
AC AAC64621;
XX
DT 26-FEB-2001 (first entry)
XX
DE Leishmania major PT3 epitope DNA sequence SEQ ID NO:9.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Leishmania major.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36345.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 48 BP; 12 A; 11 C; 11 G; 14 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.4e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48

RESULT 2
AAC64609/c
ID AAC64609 standard; DNA; 78 BP.
XX
AC AAC64609;
XX
DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)
XX
DE AgfA (SEF17) recombinant agfA::PT3a generating PCR primer 17-B.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.
XX
OS Salmonella enteritidis.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Example 2; Page 61; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 78 BP; 17 A; 17 C; 21 G; 23 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 9.5e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 48 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 1

RESULT 3
AAC64610
ID AAC64610 standard; DNA; 78 BP.
XX
AC AAC64610;
XX
DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)
XX
DE AgfA (SEF17) recombinant agfA::PT3a generating PCR primer 17-C.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.
XX
OS Salmonella enteritidis.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Example 2; Page 61; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 78 BP; 21 A; 19 C; 17 G; 21 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 9.5e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48

RESULT 4
AAC64628
ID AAC64628 standard; DNA; 456 BP.
XX
AC AAC64628;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#7 DNA sequence SEQ ID NO:23.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36352.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 331 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 378

RESULT 5
AAC64622
ID AAC64622 standard; DNA; 456 BP.
XX
AC AAC64622;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#1 DNA sequence SEQ ID NO:11.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36346.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX

PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 382 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 429

RESULT 6
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 196 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243

RESULT 7
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#9 DNA sequence SEQ ID NO:27.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.

```
DR   P-PSDB; AAB36354.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 48;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-09;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 288


RESULT 8
AAC64629
ID   AAC64629 standard; DNA; 456 BP.
XX
AC   AAC64629;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36353.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 138; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 48;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-09;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 219


RESULT 9
AAC64625
ID   AAC64625 standard; DNA; 456 BP.
XX
AC   AAC64625;
XX
DT   26-FEB-2001  (first entry)
XX
DE   AgfA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
```

XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 127 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 174

RESULT 10
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#10 DNA sequence SEQ ID NO:29.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36355.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 307 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 354

RESULT 11
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#2 DNA sequence SEQ ID NO:13.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.

XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 352 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 399

RESULT 12
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36351.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||||||||||||||||||||||||||||||||||||||||||||||
Db 262 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 309

RESULT 13
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; ds.

```
XX
OS   Salmonella enteritidis.
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
DR   P-PSDB; AAB36348.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Page 136; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention
XX
SQ   Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 48;  DB 3;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-09;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 114


RESULT 14
AAC64616
ID   AAC64616 standard; DNA; 48 BP.
XX
AC   AAC64616;
XX
DT   06-AUG-2003  (revised)
DT   26-FEB-2001  (first entry)
XX
DE   S. enteritidis recombinant agfA gene sequence.
XX
KW   Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW   vaccine; immune response; immunogen; ds.
XX
OS   Salmonella enteritidis.
OS   Synthetic.
XX
PN   WO200060102-A2.
XX
PD   12-OCT-2000.
XX
PF   05-APR-2000; 2000WO-CA000356.
XX
PR   05-APR-1999;   99US-0127888P.
XX
PA   (UYVI-) UNIV VICTORIA.
XX
PI   White AP,  Doran JL,  Collison SK,  Kay WW;
XX
DR   WPI; 2000-672631/65.
XX
PT   Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT   which encodes foreign epitope or antigen, expresses recombinant AgfA
PT   protein useful for eliciting immune response in animal.
XX
PS   Disclosure; Fig 1c; 139pp; English.
XX
CC   The present invention describes a recombinant agfA gene (I) where a
CC   segment of the gene has been replaced by a segment of a foreign DNA
CC   sequence which encodes a foreign epitope or antigen. Also described are:
CC   (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC   assembly system of strains of Salmonella, Escherichia coli and
CC   Enterobacteriaceae for the production of fimbriae comprising recombinant
CC   AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC   directing recombination of a recombinant gene into the chromosome of the
CC   homologous species; (3) directing recombination of a recombinant gene
CC   back into the chromosome of the homologous species, replacing the native
CC   copy of that gene; and (4) eliciting an immune response in an animal,
CC   comprising separating an amino acid polymer comprising a recombinant AgfA
CC   protein containing a replacement segment or segments of foreign amino
CC   acid sequence or sequences grown on a Salmonella, E. coli or
CC   Enterobacteriaceae host cell, from the host cell and introducing the
CC   polymer into the animal in conjunction with a carrier or diluent. (I) is
CC   useful for the expression of recombinant AgfA protein which is useful for
CC   eliciting an immune response in an animal. In a fimbrial presentation
CC   system the heterologous antigens are presented in high numbers (up to
CC   500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC   immunogenicity and adhesion properties relevant for an efficient live
CC   vaccine, the carrier fimbrial subunit proteins are usually strong
CC   immunogens, which may be important for directing an immune response
CC   against the inserted epitope, and hybrid fimbriae are easy and
CC   inexpensive to purify in large amount. The present sequence is given in
CC   the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC   correct OS field.)
XX
SQ   Sequence 48 BP; 14 A; 10 C; 10 G; 14 T; 0 U; 0 Other;

  Query Match             90.0%;  Score 43.2;  DB 3;  Length 48;
  Best Local Similarity   93.8%;  Pred. No. 7.6e-08;
  Matches   45;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||||||||||| ||||| |||||||||||| ||||||||||||||||||
Db          1 TATGATCAGCTCGTTACACGTGTTGTTACACATGAAATGGCACATGCA 48


RESULT 15
AAC64606
ID   AAC64606 standard; DNA; 78 BP.
XX
AC   AAC64606;
XX
```

DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)
XX
DE SefA (SEF14) recombinant sefA::PT3 generating PCR primer 14-C.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.
XX
OS Salmonella enteritidis.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Example 2; Page 61; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 78 BP; 16 A; 16 C; 21 G; 25 T; 0 U; 0 Other;

Query Match 84.6%; Score 40.6; DB 3; Length 78;
Best Local Similarity 91.5%; Pred. No. 1e-06;
Matches 43; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              ||||||||||| ||||| |||||||||||| |||||||||||| |||||
Db          1 TATGATCAGCTCGTTACACGTGTTGTTACACATGAAATGGCTCATGC 47
```

Search completed: March 18, 2004, 09:11:54
Job time : 32.3584 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:05:12 ; Search time 5.52601 Seconds
(without alignments)
4820.410 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	24.6	51.3	3636	4	US-08-961-527-78	Sequence 78, Appl
	2	22.4	46.7	1014	4	US-09-325-932A-15	Sequence 15, Appl
c	3	22	45.8	5406	4	US-08-961-527-166	Sequence 166, App
c	4	21.8	45.4	7754	4	US-09-634-238-3	Sequence 3, Appli
c	5	21.4	44.6	648	4	US-09-252-991A-11586	Sequence 11586, A
c	6	21.4	44.6	963	4	US-09-252-991A-11645	Sequence 11645, A
	7	21.4	44.6	987	4	US-09-252-991A-11794	Sequence 11794, A
c	8	21.4	44.6	1983	4	US-09-453-702B-36	Sequence 36, Appl
	9	21.4	44.6	2166	2	US-08-317-401E-3	Sequence 3, Appli
	10	21.4	44.6	2877	2	US-08-317-401E-1	Sequence 1, Appli
c	11	21.4	44.6	9541	4	US-08-961-527-132	Sequence 132, App
	12	21	43.8	597	4	US-09-543-681A-3167	Sequence 3167, Ap
	13	21	43.8	2975	4	US-08-879-337-10	Sequence 10, Appl
	14	21	43.8	6587	4	US-08-879-337-3	Sequence 3, Appli
	15	21	43.8	100848	4	US-09-596-002-39	Sequence 39, Appl
c	16	20.8	43.3	705	4	US-09-543-681A-1102	Sequence 1102, Ap
	17	20.8	43.3	1029	4	US-09-328-352-2205	Sequence 2205, Ap
	18	20.6	42.9	1239	4	US-09-328-352-779	Sequence 779, App
	19	20.6	42.9	1872	4	US-09-280-116-110	Sequence 110, App
	20	20.6	42.9	3366	4	US-09-345-650-2	Sequence 2, Appli
	21	20.6	42.9	8334	4	US-09-543-681A-1952	Sequence 1952, Ap
	22	20.4	42.5	440	4	US-09-621-976-8341	Sequence 8341, Ap
	23	20.4	42.5	753	4	US-09-252-991A-13584	Sequence 13584, A
c	24	20.4	42.5	765	4	US-09-252-991A-13860	Sequence 13860, A
	25	20.4	42.5	1404	4	US-09-107-532A-2521	Sequence 2521, Ap
	26	20.4	42.5	1608	4	US-09-328-352-2416	Sequence 2416, Ap
	27	20.4	42.5	1848	4	US-08-637-670-23	Sequence 23, Appl
c	28	20.2	42.1	1203	4	US-09-107-532A-2612	Sequence 2612, Ap
c	29	20	41.7	562	3	US-09-328-111-468	Sequence 468, App
	30	20	41.7	1152	4	US-09-489-039A-2755	Sequence 2755, Ap
	31	20	41.7	1280	3	US-08-335-844A-10	Sequence 10, Appl
	32	20	41.7	1280	4	US-09-129-366-10	Sequence 10, Appl
	33	20	41.7	1293	3	US-08-335-844A-11	Sequence 11, Appl
	34	20	41.7	1293	4	US-09-129-366-11	Sequence 11, Appl
	35	20	41.7	2472	3	US-08-335-844A-7	Sequence 7, Appli
	36	20	41.7	2472	4	US-09-129-366-7	Sequence 7, Appli
	37	20	41.7	3213	4	US-09-107-532A-62	Sequence 62, Appl
	38	20	41.7	3296	3	US-08-335-844A-14	Sequence 14, Appl
	39	20	41.7	3296	4	US-09-129-366-14	Sequence 14, Appl
	40	20	41.7	3305	3	US-08-335-844A-8	Sequence 8, Appli
	41	20	41.7	3305	4	US-09-129-366-8	Sequence 8, Appli
	42	20	41.7	3319	3	US-08-335-844A-15	Sequence 15, Appl
	43	20	41.7	3319	4	US-09-129-366-15	Sequence 15, Appl
	44	20	41.7	3358	3	US-08-335-844A-20	Sequence 20, Appl
	45	20	41.7	3358	4	US-09-129-366-20	Sequence 20, Appl

ALIGNMENTS

```
RESULT 1
US-08-961-527-78/c
; Sequence 78, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  78:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-78

  Query Match             51.3%;  Score 24.6;  DB 4;  Length 3636;
  Best Local Similarity   76.9%;  Pred. No. 1.1;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              ||| |  ||| ||| ||||||  |||||||| |||| ||
Db       2806 TCAACAAGTTGCCCTTGTTGTGGCCCATGAATTGGCTCA 2768
```

```
RESULT 2
US-09-325-932A-15
; Sequence 15, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
;  APPLICANT:  Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION:  Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of forestry plant develc
;  FILE REFERENCE:  1022
;  CURRENT APPLICATION NUMBER: US/09/325,932A
;  CURRENT FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  206
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 15
;   LENGTH: 1014
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-09-325-932A-15

  Query Match             46.7%;  Score 22.4;  DB 4;  Length 1014;
  Best Local Similarity   66.7%;  Pred. No. 6.4;
  Matches   32;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||| |||| || | | | ||  ||| | | || | || || ||| |
Db        439 TATGACCAGCAGGGTTCTCTTGCAGTTTCACCTGCAGTGACAGATGGA 486


RESULT 3
US-08-961-527-166/c
; Sequence 166, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  166:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5406 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-166

  Query Match             45.8%;  Score 22;  DB 4;  Length 5406;
  Best Local Similarity   67.4%;  Pred. No. 15;
  Matches   31;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | ||||||||     ||| | |  |||| ||| | |||| | ||||
Db       3442 AAGATCAGCTAACAGCCCTTATGATTACTCATCATATGGAAGATGC 3397


RESULT 4
US-09-634-238-3/c
; Sequence 3, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
;  APPLICANT:  Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka J.
;  APPLICANT:  Bloksberg, Leonard, N.
;  APPLICANT:  Lubbers, Mark W.
;  APPLICANT:  Dekker, James
;  APPLICANT:  Christensson, Anna C.
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul W.
;  APPLICANT:  Reid, Julian R.
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION:  Polynucleotides, materials incorporating
;  TITLE OF INVENTION:  them and methods for using them.
;  FILE REFERENCE:  11000.1043U1
;  CURRENT APPLICATION NUMBER: US/09/634,238
;  CURRENT FILING DATE:  2000-08-08
;  NUMBER OF SEQ ID NOS:  422
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 7754
;   TYPE: DNA
;   ORGANISM: Lactobacillus rhamnosus
US-09-634-238-3

  Query Match             45.4%;  Score 21.8;  DB 4;  Length 7754;
  Best Local Similarity   70.7%;  Pred. No. 20;
  Matches   29;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          5 ATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | |||| |||| |    ||| || | |||||| |||| |||
Db       4023 AGCAGCGGGTTGCTACCGTTATTGCGCATGAACTGGCGCAT 3983


RESULT 5
US-09-252-991A-11586/c
; Sequence 11586, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11586
;   LENGTH: 648
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11586

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 648;
  Best Local Similarity   71.8%;  Pred. No. 14;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              ||||  || | | |  |||||||| |||||| | ||| |
```

```
Db        296 GATCGCCTCGATCCTGGTGTTGTTTCCCATGCACTGGAA 258


RESULT 6
US-09-252-991A-11645/c
; Sequence 11645, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11645
;   LENGTH: 963
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11645

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 963;
  Best Local Similarity   71.8%;  Pred. No. 16;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              ||||  || | | |  |||||||| |||||| | ||| |
Db        316 GATCGCCTCGATCCTGGTGTTGTTTCCCATGCACTGGAA 278


RESULT 7
US-09-252-991A-11794
; Sequence 11794, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11794
;   LENGTH: 987
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11794

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 987;
  Best Local Similarity   71.8%;  Pred. No. 16;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              ||||  || | | |  |||||||| |||||| | ||| |
Db        681 GATCGCCTCGATCCTGGTGTTGTTTCCCATGCACTGGAA 719


RESULT 8
US-09-453-702B-36/c
; Sequence 36, Application US/09453702B
; Patent No. 6365723
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/453,702B
;             FILING DATE: 03-Dec-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 36:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1983
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-453-702B-36

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 1983;
  Best Local Similarity   66.0%;  Pred. No. 20;
  Matches   31;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||||||  |||     | | | |||||||| ||||||   | ||| |
Db        372 ATGATCAATTGGACCAACTTTTGGTTACCCGTGAAATTCAAGATGAA 326


RESULT 9
US-08-317-401E-3
; Sequence 3, Application US/08317401E
; Patent No. 5922561
;  GENERAL INFORMATION:
;    APPLICANT:  Thompson, Sheryl Ann
;    APPLICANT:  Yaver, Debbie Sue
;    TITLE OF INVENTION:  GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
;    TITLE OF INVENTION:  ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,401E
```

```
;      FILING DATE:  03-October-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4248.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2166 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
US-08-317-401E-3

  Query Match             44.6%;  Score 21.4;  DB 2;  Length 2166;
  Best Local Similarity   80.6%;  Pred. No. 20;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATG 34
              ||| |||||||| |||  |||| | ||||||
Db        326 GATGAGCTGGTTTCCCTGGTTGATCCCCATG 356


RESULT 10
US-08-317-401E-1
; Sequence 1, Application US/08317401E
; Patent No. 5922561
;  GENERAL INFORMATION:
;    APPLICANT:  Thompson, Sheryl Ann
;    APPLICANT:  Yaver, Debbie Sue
;    TITLE OF INVENTION:  GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
;    TITLE OF INVENTION:  ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,401E
;      FILING DATE:  03-October-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4248.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2877 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join (126..203, 253..1776)
US-08-317-401E-1

  Query Match             44.6%;  Score 21.4;  DB 2;  Length 2877;
  Best Local Similarity   80.6%;  Pred. No. 22;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATG 34
              ||| |||||||| |||  |||| | ||||||
Db        418 GATGAGCTGGTTTCCCTGGTTGATCCCCATG 448


RESULT 11
US-08-961-527-132/c
; Sequence 132, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  132:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9541 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-132

  Query Match             44.6%;  Score 21.4;  DB 4;  Length 9541;
  Best Local Similarity   71.8%;  Pred. No. 31;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          7 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              |:| || ||| ||||| | || | || |||||| | ||
Db       6099 CAGTTGATTAGCCGTGGGGCTATCAATAAAATGGGAAAT 6061


RESULT 12
US-09-543-681A-3167
; Sequence 3167, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
```

```
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 3167
;    LENGTH: 597
;    TYPE: DNA
;    ORGANISM: Proteus mirabilis
US-09-543-681A-3167

  Query Match             43.8%;  Score 21;  DB 4;  Length 597;
  Best Local Similarity   73.0%;  Pred. No. 21;
  Matches   27;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         12 GGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||     ||||| ||| | ||||||| |||| |||
Db        375 GGTTGGAACTGTTGCTACACTTGAAATGACACAAGCA 411


RESULT 13
US-08-879-337-10
; Sequence 10, Application US/08879337B
; Patent No. 6639130
; GENERAL INFORMATION:
;  APPLICANT: Jang, Jyan-Chyun
;  APPLICANT: Sheen, Jen
;  TITLE OF INVENTION: PLANT STEROL REDUCTASES AND USES THEREOF
;  FILE REFERENCE: 00786/338001
;  CURRENT APPLICATION NUMBER: US/08/879,337B
;  CURRENT FILING DATE: 1997-06-20
;  EARLIER APPLICATION NUMBER: 60/022,086
;  EARLIER FILING DATE: 1996-06-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;    LENGTH: 2975
;    TYPE: DNA
;    ORGANISM: Arabidopsis thaliana
;    FEATURE:
;    NAME/KEY: variation
;    LOCATION: (1)...(2975)
;    OTHER INFORMATION: N can be any nucleotide.
US-08-879-337-10

  Query Match             43.8%;  Score 21;  DB 4;  Length 2975;
  Best Local Similarity   66.7%;  Pred. No. 32;
  Matches   30;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | | ||||| ||| |  |    | |||||| ||| ||||| | ||
Db        310 TTTTATCAGATGGCTCTCAACTTCGTTACCGATGCAATGGTATAT 354


RESULT 14
US-08-879-337-3
; Sequence 3, Application US/08879337B
; Patent No. 6639130
; GENERAL INFORMATION:
;  APPLICANT: Jang, Jyan-Chyun
;  APPLICANT: Sheen, Jen
;  TITLE OF INVENTION: PLANT STEROL REDUCTASES AND USES THEREOF
;  FILE REFERENCE: 00786/338001
;  CURRENT APPLICATION NUMBER: US/08/879,337B
;  CURRENT FILING DATE: 1997-06-20
;  EARLIER APPLICATION NUMBER: 60/022,086
;  EARLIER FILING DATE: 1996-06-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;    LENGTH: 6587
;    TYPE: DNA
;    ORGANISM: Arabidopsis thaliana
;    FEATURE:
;    NAME/KEY: variation
;    LOCATION: (1)...(6587)
;    OTHER INFORMATION: N can be any nucleotide.
US-08-879-337-3

  Query Match             43.8%;  Score 21;  DB 4;  Length 6587;
  Best Local Similarity   66.7%;  Pred. No. 40;
  Matches   30;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | | ||||| ||| |  |    | |||||| ||| ||||| | ||
Db       3664 TTTTATCAGATGGCTCTCAACTTCGTTACCGATGCAATGGTATAT 3708


RESULT 15
US-09-596-002-39
; Sequence 39, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;  APPLICANT:      Lagace, Robert, E.
;  APPLICANT:  Patterson, Chandra
;  APPLICANT:  Berg, Kim, L.
;  TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;  FILE REFERENCE:  PM-0008-4 US
;  CURRENT APPLICATION NUMBER: US/09/596,002
;  CURRENT FILING DATE:  2000-06-16
;  PRIOR APPLICATION NUMBER:  60/140,121
;  PRIOR FILING DATE:  1999-06-18
;  NUMBER OF SEQ ID NOS:      41
;  SOFTWARE:      PERL Program
; SEQ ID NO 39
;    LENGTH: 100848
;    TYPE: DNA
;    ORGANISM: M. catarrhalis
;    FEATURE:
;    NAME/KEY: misc_feature
;    OTHER INFORMATION: Incyte template ID No. 6632636 39
;  PUBLICATION INFORMATION:
US-09-596-002-39

  Query Match             43.8%;  Score 21;  DB 4;  Length 100848;
  Best Local Similarity   66.7%;  Pred. No. 86;
  Matches   30;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||  |||  | || | |    ||| |||| |||||||||| | |
Db      73642 GATGTGCTATTGACACATACATTTATCCATCAAATGGCACAAGAA 73686


Search completed: March 18, 2004, 20:20:10
Job time : 6.52601 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 15:02:59 ; Search time 26.3931 Seconds
(without alignments)
6727.501 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	27.8	57.9	2547	12	US-10-282-122A-36154	Sequence 36154, A
	2	24.6	51.3	2547	12	US-10-282-122A-37710	Sequence 37710, A
c	3	24.6	51.3	3636	12	US-10-158-844-78	Sequence 78, Appl
c	4	24	50.0	409	12	US-10-424-599-119735	Sequence 119735,
	5	23.2	48.3	720	12	US-10-425-114-6010	Sequence 6010, Ap
	6	23	47.9	420	15	US-10-398-221-318	Sequence 318, App
	7	23	47.9	420	15	US-10-398-221-2306	Sequence 2306, Ap
c	8	23	47.9	436	10	US-09-918-995-26274	Sequence 26274, A
c	9	23	47.9	1163020	15	US-10-398-221-10	Sequence 10, Appl
c	10	23	47.9	3011208	15	US-10-398-221-2058	Sequence 2058, Ap
	11	22.6	47.1	1767	12	US-10-282-122A-25322	Sequence 25322, A
	12	22.6	47.1	3984	15	US-10-369-493-42554	Sequence 42554, A
	13	22.4	46.7	1014	14	US-10-219-220-15	Sequence 15, Appl
	14	22.4	46.7	1925	14	US-10-219-220-217	Sequence 217, App
	15	22.2	46.2	912	9	US-09-738-626-2582	Sequence 2582, Ap
c	16	22.2	46.2	2370	15	US-10-369-493-32121	Sequence 32121, A
c	17	22.2	46.2	2487	15	US-10-369-493-44598	Sequence 44598, A
c	18	22.2	46.2	2574	15	US-10-369-493-34295	Sequence 34295, A
c	19	22.2	46.2	113604	14	US-10-227-195A-1	Sequence 1, Appli
c	20	22.2	46.2	113604	14	US-10-227-195A-2	Sequence 2, Appli
c	21	22.2	46.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
c	22	22	45.8	145	9	US-09-864-761-24629	Sequence 24629, A
c	23	22	45.8	480	14	US-10-226-956-320	Sequence 320, App
c	24	22	45.8	557	9	US-09-864-761-7925	Sequence 7925, Ap
c	25	22	45.8	1791	9	US-09-822-830A-241	Sequence 241, App
c	26	22	45.8	5406	12	US-10-158-844-166	Sequence 166, App
	27	21.8	45.4	1470	12	US-10-424-599-56999	Sequence 56999, A
	28	21.8	45.4	1510	12	US-10-425-114-36154	Sequence 36154, A
	29	21.8	45.4	1646	12	US-10-425-114-31422	Sequence 31422, A
	30	21.8	45.4	2394	12	US-10-425-114-32979	Sequence 32979, A
	31	21.8	45.4	3328	12	US-10-425-114-34623	Sequence 34623, A
c	32	21.8	45.4	7754	15	US-10-264-213-3	Sequence 3, Appli
c	33	21.8	45.4	8495	13	US-10-108-605-356	Sequence 356, App
c	34	21.6	45.0	633	9	US-09-917-800A-1050	Sequence 1050, Ap
	35	21.6	45.0	729	9	US-09-738-626-1352	Sequence 1352, Ap
	36	21.6	45.0	852	12	US-10-627-476-367	Sequence 367, App
c	37	21.6	45.0	1580	12	US-10-425-114-21821	Sequence 21821, A
	38	21.6	45.0	1746	15	US-10-260-238-578	Sequence 578, App
	39	21.6	45.0	2853	12	US-10-425-114-13682	Sequence 13682, A
c	40	21.6	45.0	1691139	14	US-10-067-514-1	Sequence 1, Appli
c	41	21.6	45.0	1691139	15	US-10-419-723-1	Sequence 1, Appli
	42	21.6	45.0	3309400	9	US-09-738-626-1	Sequence 1, Appli
	43	21.4	44.6	1011	12	US-10-282-122A-30293	Sequence 30293, A
	44	21.4	44.6	1179	10	US-09-769-744A-119	Sequence 119, App
c	45	21.4	44.6	1983	14	US-10-114-170-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-10-282-122A-36154
; Sequence 36154, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

```
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36154
;   LENGTH: 2547
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
US-10-282-122A-36154

  Query Match             57.9%;  Score 27.8;  DB 12;  Length 2547;
  Best Local Similarity   82.1%;  Pred. No. 0.27;
  Matches   32;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          7 CAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              |||| |||| | | ||||||| |||||||| |||| |||
Db        850 CAGCAGGTTGCACTTGTTGTTGCCCATGAACTGGCGCAT 888


RESULT 2
US-10-282-122A-37710
; Sequence 37710, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37710
;   LENGTH: 2547
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-10-282-122A-37710

  Query Match             51.3%;  Score 24.6;  DB 12;  Length 2547;
  Best Local Similarity   76.9%;  Pred. No. 6;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              ||| |  ||| ||| |||||||  |||||||| |||| ||
Db        849 TCAACAAGTTGCCCTTGTTGTGGCCCATGAATTGGCTCA 887


RESULT 3
US-10-158-844-78/c
; Sequence 78, Application US/10158844
; Publication No. US20040029118A1
;  GENERAL INFORMATION:
;       APPLICANT: Kunsch et al.
;       TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
;       NUMBER OF SEQUENCES: 391
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD-R
;            COMPUTER: Dell Latitude Pentium 3
;            OPERATING SYSTEM: Windows 98
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/158,844
;            FILING DATE: 03-Jun-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 08/961,527
;            FILING DATE: 1997-10-30
;            APPLICATION NUMBER: US 60/029,960
;            FILING DATE: 1996-10-31
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Hyman, Mark J.
;            REGISTRATION NUMBER: 46,789
;            REFERENCE/DOCKET NUMBER: PB340P1D1
;   INFORMATION FOR SEQ ID NO: 78:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 3636 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-10-158-844-78

  Query Match             51.3%;  Score 24.6;  DB 12;  Length 3636;
  Best Local Similarity   76.9%;  Pred. No. 6.6;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              ||| |  ||| ||| |||||||  |||||||| |||| ||
Db       2806 TCAACAAGTTGCCCTTGTTGTGGCCCATGAATTGGCTCA 2768


RESULT 4
US-10-424-599-119735/c
; Sequence 119735, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
```

```
; SEQ ID NO 119735
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_79129C.1
US-10-424-599-119735

  Query Match             50.0%;  Score 24;  DB 12;  Length 409;
  Best Local Similarity   68.8%;  Pred. No. 6.2;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              || | | | | | || ||||||||| |  |||  |||| ||| ||| ||
Db        340 TACGGTGACCAGCTTGCCCGTGTTCTCGCCCGAGAAAAGGCGCATCCA 293


RESULT 5
US-10-425-114-6010
; Sequence 6010, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 6010
;   LENGTH: 720
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700550972_FLI
US-10-425-114-6010

  Query Match             48.3%;  Score 23.2;  DB 12;  Length 720;
  Best Local Similarity   70.5%;  Pred. No. 16;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              ||| |||||||||| ||  | || || ||  ||| |  || |||
Db        285 ATGTTCAGCTGGTTCCCGATCTTTTTCCCATTGAGAAAGCCCAT 328


RESULT 6
US-10-398-221-318
; Sequence 318, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 318
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-318

  Query Match             47.9%;  Score 23;  DB 15;  Length 420;
  Best Local Similarity   68.1%;  Pred. No. 16;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | |||| | | |||||    ||| ||| | |||||| | |  |||||
Db        175 TCTGATTATCAGGTTAAGTATGTAGTTGCTCATGAACTAGGGCATGC 221


RESULT 7
US-10-398-221-2306
; Sequence 2306, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2306
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2306

  Query Match             47.9%;  Score 23;  DB 15;  Length 420;
  Best Local Similarity   68.1%;  Pred. No. 16;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | |||| | | |||||    ||| ||| | |||||| | |  |||||
Db        175 TCTGATTATCAGGTTAAGTATGTAGTTGCTCATGAACTAGGGCATGC 221


RESULT 8
US-09-918-995-26274/c
; Sequence 26274, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: PastSEQ for Windows Version 3.0
; SEQ ID NO 26274
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(436)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26274

  Query Match             47.9%;  Score 23;  DB 10;  Length 436;
  Best Local Similarity   68.1%;  Pred. No. 17;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              | ||||| ||||||| | | || |||  |  |||| | || ||  ||
```

```
Db        347 ACGATCACCTGGTTATCTGGGTGGTTGACAGTGAAGTAGCCCACACA 301


RESULT 9
US-10-398-221-10/c
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;   LENGTH: 1163020
;   TYPE: DNA
;   ORGANISM: Listeria innocua
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(end)
;   OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

  Query Match             47.9%;  Score 23;  DB 15;  Length 1163020;
  Best Local Similarity   68.1%;  Pred. No. 1.7e+02;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | ||||  | | |||||    ||| ||| | |||||| | |  |||||
Db    1158751 TCTGATTATCAGGTTAAGTATGTAGTTGCTCATGAACTAGGGCATGC 1158705


RESULT 10
US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
;   LENGTH: 3011208
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2058

  Query Match             47.9%;  Score 23;  DB 15;  Length 3011208;
  Best Local Similarity   68.1%;  Pred. No. 2.1e+02;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              | ||||  | | |||||    ||| ||| | |||||| | |  |||||
Db    1249014 TCTGATTATCAGGTTAAGTATGTAGTTGCTCATGAACTAGGGCATGC 1248968


RESULT 11
US-10-282-122A-25322
; Sequence 25322, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25322
;   LENGTH: 1767
;   TYPE: DNA
;   ORGANISM: Legionella pneumophila
US-10-282-122A-25322

  Query Match             47.1%;  Score 22.6;  DB 12;  Length 1767;
  Best Local Similarity   68.9%;  Pred. No. 37;
  Matches   31;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              |||   |||  ||||| | |||  |||   |||| ||| ||||||
Db        929 ATGTCGAGCGAGTTACTCATGTGATTAATTATGATATGCCACATG 973


RESULT 12
US-10-369-493-42554
; Sequence 42554, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
```

```
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42554
;   LENGTH: 3984
;   TYPE: DNA
;   ORGANISM: Anabaena PCC7120
US-10-369-493-42554

  Query Match             47.1%;  Score 22.6;  DB 15;  Length 3984;
  Best Local Similarity   75.7%;  Pred. No. 47;
  Matches   28;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              ||| || |||||| ||| || | ||||| |  |||||
Db        843 TCACCTAGTTACCGGTGATGATGCCCATTATGTGGCA 879


RESULT 13
US-10-219-220-15
; Sequence 15, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
;  APPLICANT: Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION: Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of plant development
;  FILE REFERENCE: 11000.1022c1
;  CURRENT APPLICATION NUMBER: US/10/219,220
;  CURRENT FILING DATE:  2002-08-14
;  PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
;  PRIOR FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 290
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15
;   LENGTH: 1014
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-10-219-220-15

  Query Match             46.7%;  Score 22.4;  DB 14;  Length 1014;
  Best Local Similarity   66.7%;  Pred. No. 38;
  Matches   32;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||| |||| || | | | ||  ||| | | || | || || ||| |
Db        439 TATGACCAGCAGGGTTCTCTTGCAGTTTCACCTGCAGTGACAGATGGA 486


RESULT 14
US-10-219-220-217
; Sequence 217, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
;  APPLICANT: Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION: Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of plant development
;  FILE REFERENCE: 11000.1022c1
;  CURRENT APPLICATION NUMBER: US/10/219,220
;  CURRENT FILING DATE:  2002-08-14
;  PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
;  PRIOR FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 290
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 217
;   LENGTH: 1925
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-10-219-220-217

  Query Match             46.7%;  Score 22.4;  DB 14;  Length 1925;
  Best Local Similarity   66.7%;  Pred. No. 46;
  Matches   32;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||| |||| || | | | ||  ||| | | || | || || ||| |
Db        980 TATGACCAGCAGGGTTCTCTTGCAGTTTCACCTGCAGTGACAGATGGA 1027


RESULT 15
US-09-738-626-2582
; Sequence 2582, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2582
;   LENGTH: 912
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-2582

  Query Match             46.2%;  Score 22.2;  DB 9;  Length 912;
  Best Local Similarity   69.8%;  Pred. No. 44;
  Matches   30;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          3 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              ||||  |||||| |  |||| |||   | |||| | ||||| |
Db        180 TGATGCGCTGGTCAAGCGTGGTGTCGTCAATGAGAAGGCACGT 222


Search completed: March 19, 2004, 03:49:08
Job time : 39.3931 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:28:02 ; Search time 279.029 Seconds
(without alignments)
6052.899 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	100.0	48	23	US-09-543-407-9	Sequence 9, Appli
c 2	48	100.0	78	23	US-09-543-407-52	Sequence 52, Appl
3	48	100.0	78	23	US-09-543-407-53	Sequence 53, Appl
4	48	100.0	456	23	US-09-543-407-11	Sequence 11, Appl
5	48	100.0	456	23	US-09-543-407-13	Sequence 13, Appl
6	48	100.0	456	23	US-09-543-407-15	Sequence 15, Appl
7	48	100.0	456	23	US-09-543-407-17	Sequence 17, Appl
8	48	100.0	456	23	US-09-543-407-19	Sequence 19, Appl
9	48	100.0	456	23	US-09-543-407-21	Sequence 21, Appl
10	48	100.0	456	23	US-09-543-407-23	Sequence 23, Appl
11	48	100.0	456	23	US-09-543-407-25	Sequence 25, Appl
12	48	100.0	456	23	US-09-543-407-27	Sequence 27, Appl
13	48	100.0	456	23	US-09-543-407-29	Sequence 29, Appl
c 14	40.6	84.6	78	23	US-09-543-407-48	Sequence 48, Appl
15	40.6	84.6	78	23	US-09-543-407-49	Sequence 49, Appl
16	29.4	61.2	3105	3	US-07-965-673A-9	Sequence 9, Appli
17	27.8	57.9	2547	48	US-10-282-122A-36154	Sequence 36154, A
c 18	26.8	55.8	107223	21	US-09-528-237A-850	Sequence 850, App
19	24.6	51.3	1780	1	PCT-US97-22578-44	Sequence 44, Appl
20	24.6	51.3	2448	15	US-09-107-433-2456	Sequence 2456, Ap
21	24.6	51.3	2448	53	US-10-617-320-2456	Sequence 2456, Ap
22	24.6	51.3	2544	1	PCT-US03-27401-93	Sequence 93, Appl
23	24.6	51.3	2544	52	US-10-472-928-1519	Sequence 1519, Ap
24	24.6	51.3	2547	24	US-09-583-110-77	Sequence 77, Appl
25	24.6	51.3	2547	48	US-10-282-122A-37710	Sequence 37710, A
26	24.6	51.3	2547	53	US-10-640-833-77	Sequence 77, Appl
27	24.6	51.3	2867	62	US-60-068-175-437	Sequence 437, App
c 28	24.6	51.3	3636	46	US-10-158-844-78	Sequence 78, Appl
29	24.6	51.3	3657	58	US-60-029-960-47	Sequence 47, Appl
30	24.6	51.3	23131	62	US-60-061-998-613	Sequence 613, App
31	24.6	51.3	2162598	52	US-10-472-928-4979	Sequence 4979, Ap
c 32	24	50.0	409	24	US-09-565-240-34030	Sequence 34030, A
c 33	24	50.0	409	27	US-09-654-617-77662	Sequence 77662, A
c 34	24	50.0	409	29	US-09-684-016-77662	Sequence 77662, A
c 35	24	50.0	409	51	US-10-424-599-119735	Sequence 119735,
36	24	50.0	2023	26	US-09-620-392-69478	Sequence 69478, A
c 37	24	50.0	5823	51	US-10-437-963-69258	Sequence 69258, A
c 38	24	50.0	12979	30	US-09-702-134-5750	Sequence 5750, Ap
c 39	24	50.0	12979	33	US-09-815-264-58710	Sequence 58710, A
c 40	24	50.0	13038	26	US-09-620-392-30380	Sequence 30380, A
41	23.6	49.2	390	21	US-09-521-640-44946	Sequence 44946, A
c 42	23.4	48.8	119	35	US-09-865-439A-100294	Sequence 100294,
c 43	23.4	48.8	119	76	US-60-207-458-144215	Sequence 144215,
44	23.2	48.3	310	27	US-09-654-617-237766	Sequence 237766,
45	23.2	48.3	310	29	US-09-684-016-237766	Sequence 237766,

ALIGNMENTS

```
RESULT 1
US-09-543-407-9
; Sequence 9, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 48
;   TYPE: DNA
;   ORGANISM: Leishmania  major
US-09-543-407-9

  Query Match             100.0%;  Score 48;  DB 23;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 1e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48


RESULT 2
US-09-543-407-52/c
; Sequence 52, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 52
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PCR primer
US-09-543-407-52

  Query Match             100.0%;  Score 48;  DB 23;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db         48 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 1


RESULT 3
US-09-543-407-53
; Sequence 53, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 53
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PCR primer
US-09-543-407-53

  Query Match             100.0%;  Score 48;  DB 23;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
```

```
RESULT 4
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 429


RESULT 5
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 399


RESULT 6
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 114


RESULT 7
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 174


RESULT 8
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
```

```
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 243


RESULT 9
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 309


RESULT 10
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 378


RESULT 11
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 219


RESULT 12
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 456
;   TYPE: DNA
```

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 288


RESULT 13
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
;   OTHER INFORMATION: sequence containing the replacement fragment
;   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

  Query Match             100.0%;  Score 48;  DB 23;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 354


RESULT 14
US-09-543-407-48/c
; Sequence 48, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PCR primer
US-09-543-407-48

  Query Match             84.6%;  Score 40.6;  DB 23;  Length 78;
  Best Local Similarity   91.5%;  Pred. No. 1.1e-05;
  Matches   43;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              |||||||||| ||||| |||||||||| |||||||||| |||||
Db         48 TATGATCAGCTCGTTACACGTGTTGTTACACATGAAATGGCTCATGC 2


RESULT 15
US-09-543-407-49
; Sequence 49, Application US/09543407
; GENERAL INFORMATION:
;  APPLICANT: White, Aaron P.
;  APPLICANT:  Doran, James L.
;  APPLICANT:  Collinson, S. Karen
;  APPLICANT:  Kay, William W.
;  TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;  TITLE OF INVENTION:  PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;  FILE REFERENCE: 920043.406
;  CURRENT APPLICATION NUMBER: US/09/543,407
;  CURRENT FILING DATE:  2000-04-05
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PCR primer
US-09-543-407-49

  Query Match             84.6%;  Score 40.6;  DB 23;  Length 78;
  Best Local Similarity   91.5%;  Pred. No. 1.1e-05;
  Matches   43;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              |||||||||| ||||| |||||||||| |||||||||| |||||
Db          1 TATGATCAGCTCGTTACACGTGTTGTTACACATGAAATGGCTCATGC 47


Search completed: March 19, 2004, 03:01:48
Job time : 282.029 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 09:12:08 ; Search time 5.69942 Seconds
(without alignments)
3195.690 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1021000 seqs, 189724845 residues

Total number of hits satisfying chosen parameters: 2042000

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23.2	48.3	435	6	US-10-767-795-1394	Sequence 1394, Ap
	2	23.2	48.3	539	6	US-10-767-795-1392	Sequence 1392, Ap
c	3	23.2	48.3	545	6	US-10-767-701-29396	Sequence 29396, A
	4	23.2	48.3	743	6	US-10-767-701-11493	Sequence 11493, A
	5	22.2	46.2	1035	6	US-10-781-014-167	Sequence 167, App
	6	22	45.8	808	6	US-10-779-543-5864	Sequence 5864, Ap
	7	21.6	45.0	201	6	US-10-767-471-34311	Sequence 34311, A
	8	21.6	45.0	2895	6	US-10-767-701-14458	Sequence 14458, A
	9	21.6	45.0	546025	6	US-10-767-471-10717	Sequence 10717, A
c	10	21.4	44.6	607	6	US-10-767-701-31393	Sequence 31393, A
c	11	21	43.8	26371	6	US-10-765-790-41	Sequence 41, Appl
	12	20.8	43.3	175737	6	US-10-783-271-34	Sequence 34, Appl
c	13	20.6	42.9	421	6	US-10-779-543-15761	Sequence 15761, A
	14	20.6	42.9	905	6	US-10-781-469-49	Sequence 49, Appl
	15	20.6	42.9	198161	6	US-10-775-169-52	Sequence 52, Appl
c	16	20.4	42.5	201	6	US-10-767-471-11899	Sequence 11899, A
	17	20.4	42.5	1404	6	US-10-417-884A-2521	Sequence 2521, Ap
c	18	20.4	42.5	318752	6	US-10-767-471-10553	Sequence 10553, A
c	19	20.2	42.1	532	6	US-10-767-701-30765	Sequence 30765, A
	20	20.2	42.1	779	6	US-10-781-014-169	Sequence 169, App
c	21	20.2	42.1	1203	6	US-10-417-884A-2612	Sequence 2612, Ap
	22	20.2	42.1	4503	6	US-10-417-375A-31	Sequence 31, Appl
	23	20.2	42.1	106776	6	US-10-417-375A-30	Sequence 30, Appl
c	24	20	41.7	596	6	US-10-767-701-657	Sequence 657, App
	25	20	41.7	746	6	US-10-779-543-5903	Sequence 5903, Ap
c	26	20	41.7	863	6	US-10-767-701-8642	Sequence 8642, Ap
	27	20	41.7	3213	6	US-10-417-884A-62	Sequence 62, Appl
	28	20	41.7	36135	6	US-10-100-683-11827	Sequence 11827, A
	29	19.8	41.2	390	6	US-10-779-543-12187	Sequence 12187, A
c	30	19.8	41.2	594	6	US-10-417-884A-380	Sequence 380, App
	31	19.8	41.2	600	7	US-60-545-213-391	Sequence 391, App
	32	19.8	41.2	600	7	US-60-545-213-392	Sequence 392, App
	33	19.8	41.2	600	7	US-60-545-213-542	Sequence 542, App
	34	19.8	41.2	600	7	US-60-545-213-4663	Sequence 4663, Ap
	35	19.8	41.2	600	7	US-60-545-213-4664	Sequence 4664, Ap
	36	19.8	41.2	600	7	US-60-545-213-4814	Sequence 4814, Ap
c	37	19.8	41.2	757	6	US-10-100-683-5410	Sequence 5410, Ap
	38	19.8	41.2	844	6	US-10-767-701-10879	Sequence 10879, A
	39	19.8	41.2	7652	1	PCT-US03-31941A-1	Sequence 1, Appli
c	40	19.8	41.2	8684	1	PCT-US03-31941A-2	Sequence 2, Appli
c	41	19.8	41.2	10011	1	PCT-US03-31941A-19	Sequence 19, Appl
	42	19.6	40.8	201	6	US-10-767-471-34488	Sequence 34488, A
	43	19.6	40.8	201	6	US-10-767-471-34490	Sequence 34490, A
	44	19.6	40.8	201	6	US-10-767-471-34556	Sequence 34556, A
	45	19.6	40.8	1013	6	US-10-767-701-12512	Sequence 12512, A

ALIGNMENTS

RESULT 1
US-10-767-795-1394
; Sequence 1394, Application US/10767795
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 1394
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: uC-gsronu33B172b05b1
US-10-767-795-1394

Query Match 48.3%; Score 23.2; DB 6; Length 435;
Best Local Similarity 70.5%; Pred. No. 1.7;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy          3 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              |||| | ||  | || ||| | |||||| | |||   |||||||
Db         73 TGATAAACTCATAACTCGTTTGGTTACCAAGGAAGAAGCACATG 116
```

RESULT 2
US-10-767-795-1392
; Sequence 1392, Application US/10767795
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 1392
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: uC-gsronu33B027f06b1

US-10-767-795-1392

Query Match 48.3%; Score 23.2; DB 6; Length 539;
Best Local Similarity 70.5%; Pred. No. 1.7;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 3 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
Db 29 TGATAAACTCATAACTCGTTTGGTTACCAAGGAAGAAGCACATG 72

RESULT 3
US-10-767-701-29396/c
; Sequence 29396, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 29396
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 9299890
US-10-767-701-29396

Query Match 48.3%; Score 23.2; DB 6; Length 545;
Best Local Similarity 77.8%; Pred. No. 1.7;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 ATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGG 40
Db 532 ATCAGCTGGTGACCCATGCAGATACCCATGACGGGG 497

RESULT 4
US-10-767-701-11493
; Sequence 11493, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 11493
; LENGTH: 743
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS45512_1
US-10-767-701-11493

Query Match 48.3%; Score 23.2; DB 6; Length 743;
Best Local Similarity 70.5%; Pred. No. 1.9;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
Db 269 ATGTTTAGCTGGTTCCCCATCTTCTTCCCATTGAGAAAGCCCAT 312

RESULT 5
US-10-781-014-167

; Sequence 167, Application US/10781014
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN CARBON METABOLISM AND ENERGY
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: BGI-126CPCN
; CURRENT APPLICATION NUMBER: US/10/781,014
; CURRENT FILING DATE: 2004-02-17
; PRIOR APPLICATION NUMBER: US 09/602,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/143,208
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/151,572
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19931412.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931413.6
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931424.1
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931428.4
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 167
; LENGTH: 1035
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1012)
; OTHER INFORMATION: RXN02920
US-10-781-014-167

Query Match 46.2%; Score 22.2; DB 6; Length 1035;
Best Local Similarity 69.8%; Pred. No. 5.2;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 3 TGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
Db 280 TGATGCGCTGGTCAAGCGTGGTGTCGTCAATGAGAAGGCACGT 322

RESULT 6
US-10-779-543-5864
; Sequence 5864, Application US/10779543
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234

```
;  PRIOR FILING DATE: 1998-10-21
;  PRIOR APPLICATION NUMBER: 09/297,648
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: PCT/US99/01619
;  PRIOR FILING DATE: 1999-01-28
;  PRIOR APPLICATION NUMBER: 60/072,910
;  PRIOR FILING DATE: 1998-01-28
;  PRIOR APPLICATION NUMBER: 60/075,954
;  PRIOR FILING DATE: 1998-02-24
;  PRIOR APPLICATION NUMBER: 60/080,114
;  PRIOR FILING DATE: 1998-03-31
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 23767
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5864
;   LENGTH: 808
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 16, 20, 57, 68, 77, 78, 80, 89, 105, 107, 108, 115, 117,
;   LOCATION: 120, 123, 129, 133, 138, 140, 143, 147, 148, 156, 159, 163,
;   LOCATION: 166, 172, 173, 174, 180, 186, 192, 197, 198, 201, 207, 209,
;   LOCATION: 214, 216, 217, 218, 219, 226, 228, 236, 237, 238, 243
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 244, 245, 246, 254, 283, 308, 311, 317, 325, 330, 332, 339,
;   LOCATION: 357, 358, 375, 385, 398, 401, 415, 422, 424, 426, 430, 434,
;   LOCATION: 442, 448, 449, 451, 470, 471, 475, 476, 477, 479, 480, 491,
;   LOCATION: 497, 507, 510, 527, 559, 574, 576, 584, 591, 609, 636
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 639, 641, 643, 650, 660, 672, 673, 681, 684, 722, 743, 754,
;   LOCATION: 762, 783, 784, 795, 801
;   OTHER INFORMATION: n = A,T,C or G
US-10-779-543-5864

  Query Match             45.8%;  Score 22;  DB 6;  Length 808;
  Best Local Similarity   68.3%;  Pred. No. 5.9;
  Matches   28;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGC 41
              ||||||   |||   ||| ||| |  ||||||||||   ||
Db        333 TATGATNTTCTGCCCACCTGTGATNNTACCCATGAACACGC 373


RESULT 7
US-10-767-471-34311
; Sequence 34311, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34311
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-767-471-34311

  Query Match             45.0%;  Score 21.6;  DB 6;  Length 201;
  Best Local Similarity   68.2%;  Pred. No. 6.4;
  Matches   30;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              | ||||||| |   ||  | ||| | |||  |||||| || |||
Db        155 TTTGATCAGTTTCATATTCCTGTGGGTACTAATGAAAAGGAACA 198


RESULT 8
US-10-767-701-14458
; Sequence 14458, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14458
;   LENGTH: 2895
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS8018_1
US-10-767-701-14458

  Query Match             45.0%;  Score 21.6;  DB 6;  Length 2895;
  Best Local Similarity   68.2%;  Pred. No. 12;
  Matches   30;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              || | || | |||| |   ||||||  | ||||||| | || |||
Db       1033 ATAAGCAACGGGTTGCAGTTGTTGTAGCTCATGAATTAGCTCAT 1076


RESULT 9
US-10-767-471-10717
; Sequence 10717, Application US/10767471
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001505
;  CURRENT APPLICATION NUMBER: US/10/767,471
;  CURRENT FILING DATE:  2004-01-30
;  NUMBER OF SEQ ID NOS: 50231
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10717
;   LENGTH: 546025
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(546025)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-767-471-10717

  Query Match             45.0%;  Score 21.6;  DB 6;  Length 546025;
  Best Local Similarity   68.2%;  Pred. No. 36;
  Matches   30;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              | ||||||| |   ||  | ||| | |||  |||||| || |||
Db     329443 TTTGATCAGTTTCATATTCCTGTGGGTACTAATGAAAAGGAACA 329486


RESULT 10
US-10-767-701-31393/c
; Sequence 31393, Application US/10767701
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
```

```
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 31393
;   LENGTH: 607
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(607)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 18067944
US-10-767-701-31393

  Query Match             44.6%;  Score 21.4;  DB 6;  Length 607;
  Best Local Similarity   66.0%;  Pred. No. 9.8;
  Matches   31;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              | |||  |||||| | | || |    || ||| | ||||||||  ||
Db        306 AGGATGGGCTGGTGAGCTGTTTATGCACGCATTACATGGCACAAACA 260


RESULT 11
US-10-765-790-41/c
; Sequence 41, Application US/10765790
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
;   LENGTH: 26371
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-41

  Query Match             43.8%;  Score 21;  DB 6;  Length 26371;
  Best Local Similarity   66.7%;  Pred. No. 33;
  Matches   30;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              | |  |||||||| ||  | ||   | ||||  |||| | |||||
Db      15911 AAGCCCAGCTGGTGACTTGAGTGTCTCCCCAACAAATAGAACATG 15867


RESULT 12
US-10-783-271-34
; Sequence 34, Application US/10783271
; GENERAL INFORMATION:
;  APPLICANT: Veridex, LLC
;  APPLICANT:  Wang, Yixin
;  TITLE OF INVENTION: BREAST CANCER PROGNOSTICS
;  FILE REFERENCE: VDX-5003 USNP
;  CURRENT APPLICATION NUMBER: US/10/783,271
;  CURRENT FILING DATE:  2004-02-20
;  NUMBER OF SEQ ID NOS: 111
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
;   LENGTH: 175737
;   TYPE: DNA
;   ORGANISM: human
US-10-783-271-34

  Query Match             43.3%;  Score 20.8;  DB 6;  Length 175737;
  Best Local Similarity   70.0%;  Pred. No. 60;
  Matches   28;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          9 GCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||| | | | | |   || ||| |||||||||||| | ||
Db     116051 GCTGCTCATCTGGGACTTTGCCCCTGAAATGGCACCTCCA 116090


RESULT 13
US-10-779-543-15761/c
; Sequence 15761, Application US/10779543
; GENERAL INFORMATION:
;  APPLICANT: Williams et al
;  TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
;  TITLE OF INVENTION:  IN CANCEROUS CELLS AND THEIR METHODS OF USE II
;  FILE REFERENCE: 2300-21302
;  CURRENT APPLICATION NUMBER: US/10/779,543
;  CURRENT FILING DATE:  2004-02-12
;  PRIOR APPLICATION NUMBER: 10/076,555
;  PRIOR FILING DATE: 2002-02-15
;  PRIOR APPLICATION NUMBER: 09/217,471
;  PRIOR FILING DATE: 1998-12-21
;  PRIOR APPLICATION NUMBER: 60/068,755
;  PRIOR FILING DATE: 1997-12-23
;  PRIOR APPLICATION NUMBER: 60/080,664
;  PRIOR FILING DATE: 1998-04-03
;  PRIOR APPLICATION NUMBER: 60/105,234
;  PRIOR FILING DATE: 1998-10-21
;  PRIOR APPLICATION NUMBER: 09/297,648
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: PCT/US99/01619
;  PRIOR FILING DATE: 1999-01-28
;  PRIOR APPLICATION NUMBER: 60/072,910
;  PRIOR FILING DATE: 1998-01-28
;  PRIOR APPLICATION NUMBER: 60/075,954
;  PRIOR FILING DATE: 1998-02-24
;  PRIOR APPLICATION NUMBER: 60/080,114
;  PRIOR FILING DATE: 1998-03-31
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 23767
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15761
;   LENGTH: 421
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-779-543-15761

  Query Match             42.9%;  Score 20.6;  DB 6;  Length 421;
  Best Local Similarity   74.3%;  Pred. No. 19;
  Matches   26;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy         11 TGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | |||| | ||   |||||||||||||   |||||
Db        225 TTGTTAACTGTCCAGTTACCCATGAAAATTCACAT 191


RESULT 14
US-10-781-469-49
; Sequence 49, Application US/10781469
; GENERAL INFORMATION:
;  APPLICANT: Reddy, Roopa
;  APPLICANT:  Guegler, Karl, J.
;  APPLICANT:  Au-Young, Janice K.
;  TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING RECEPTOR PROTEINS
;  FILE REFERENCE: PA-0014-1 CON
;  CURRENT APPLICATION NUMBER: US/10/781,469
;  CURRENT FILING DATE:  2004-02-17
;  PRIOR APPLICATION NUMBER: US 09/496,321
;  PRIOR FILING DATE: 2000-02-01
```

; PRIOR APPLICATION NUMBER: US 60/118,905
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PERL Program
; SEQ ID NO 49
; LENGTH: 905
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 2832568
US-10-781-469-49

Query Match 42.9%; Score 20.6; DB 6; Length 905;
Best Local Similarity 67.4%; Pred. No. 23;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACA 44
              || | | |  ||| ||| | ||  |  |||||||| |||| ||
Db        605 ATAAACTGTGGGTCACCAGAGTCATAGCCCATGAACTGGCGCA 647
```

RESULT 15
US-10-775-169-52
; Sequence 52, Application US/10775169
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Burczynski, Michael
; APPLICANT: Twine, Natalie
; APPLICANT: Dorner, Andrew
; APPLICANT: Trepicchio, William
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
; FILE REFERENCE: AM101080 (031896-013000)
; CURRENT APPLICATION NUMBER: US/10/775,169
; CURRENT FILING DATE: 2004-02-11
; NUMBER OF SEQ ID NOS: 5278
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 52
; LENGTH: 198161
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-169-52

Query Match 42.9%; Score 20.6; DB 6; Length 198161;
Best Local Similarity 67.4%; Pred. No. 74;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
Qy          6 TCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||  ||  | | | |||| ||| ||| |||||| | || | |
Db     175379 TCACATGTGTTCTCATGTTCTTAACCACGAAATGTCTCAGGGA 175421
```

Search completed: March 19, 2004, 03:10:06
Job time : 8.69942 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 08:00:47 ; Search time 214.058 Seconds
(without alignments)
6696.250 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaaatggcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25.2	52.5	878	28	CC001571	CC001571 PUDIE23TD
2	25	52.1	894	13	BU147350	BU147350 AGENCOURT
3	24.8	51.7	273	10	BB435111	BB435111 BB435111
c 4	24.6	51.3	1201	9	AL576676	AL576676 AL576676
5	24.2	50.4	446	9	AU299906	AU299906 AU299906
c 6	24.2	50.4	539	28	AZ056038	AZ056038 RPCI-23-4
7	24.2	50.4	903	14	CB200568	CB200568 AGENCOURT
c 8	24.2	50.4	1032	29	CNS01XVV	AL172228 Tetraodon
9	24	50.0	176	10	AW837303	AW837303 QV2-LT003
10	24	50.0	300	10	BF634308	BF634308 NF082C02D
11	24	50.0	600	12	BJ102563	BJ102563 BJ102563
c 12	24	50.0	610	9	AL589140	AL589140 AL589140
c 13	24	50.0	766	13	BU294276	BU294276 603606963
c 14	24	50.0	780	13	BU252275	BU252275 603748238
c 15	24	50.0	791	13	BU324056	BU324056 603493436
c 16	24	50.0	812	13	BU328355	BU328355 603406191
c 17	24	50.0	813	13	BU202575	BU202575 604152384
c 18	23.8	49.6	294	9	AA224363	AA224363 zr14b09.s
c 19	23.6	49.2	296	10	AW373707	AW373707 QV4-BT053
c 20	23.6	49.2	316	14	CB017485	CB017485 pgn1c.pk0
c 21	23.6	49.2	994	13	BU250654	BU250654 603321055
22	23.4	48.8	446	10	BE443880	BE443880 WHE1124_F
23	23.4	48.8	605	14	CB891544	CB891544 EST648513
c 24	23.4	48.8	673	14	CD452770	CD452770 WHE1124_F
25	23.4	48.8	1016	29	CC600714	CC600714 ZMMBBc040
c 26	23.2	48.3	153	13	BQ172708	BQ172708 1091027E1
27	23.2	48.3	230	10	BB435292	BB435292 BB435292
c 28	23.2	48.3	248	14	CF245156	CF245156 3530_1_8_
29	23.2	48.3	275	10	BB435052	BB435052 BB435052
30	23.2	48.3	295	10	BB436405	BB436405 BB436405
31	23.2	48.3	305	10	BB433896	BB433896 BB433896
c 32	23.2	48.3	460	12	BM379062	BM379062 MEST573-G
c 33	23.2	48.3	467	13	BQ744659	BQ744659 946106F04
34	23.2	48.3	479	12	BG947566	BG947566 IP1_6_B09
35	23.2	48.3	521	14	CD204019	CD204019 HS1_3_C07
36	23.2	48.3	522	12	BI076348	BI076348 IP1_26_B0
c 37	23.2	48.3	540	9	AI711846	AI711846 605064G07
c 38	23.2	48.3	545	10	BE358333	BE358333 DG1_29_D0
c 39	23.2	48.3	577	14	CA404413	CA404413 EL01N0517
40	23.2	48.3	590	12	BG103593	BG103593 RHIZ2_38_
41	23.2	48.3	591	12	BM318096	BM318096 PI1_78_D1
42	23.2	48.3	615	10	BE365335	BE365335 PI1_26_B1
c 43	23.2	48.3	620	14	CF049562	CF049562 QCL38b09.
44	23.2	48.3	623	10	BE919209	BE919209 FM1_4_E01
c 45	23.2	48.3	623	14	CD527595	CD527595 3529_1_12

ALIGNMENTS

RESULT 1
CC001571
LOCUS CC001571 878 bp DNA linear GSS 31-MAR-2003
DEFINITION PUDIE23TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa217C21, genomic survey sequence.
ACCESSION CC001571
VERSION CC001571.1 GI:29380131
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 878)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.

```
FEATURES             Location/Qualifiers
     source          1. .878
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa217C21"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             52.5%;  Score 25.2;  DB 28;  Length 878;
  Best Local Similarity   71.7%;  Pred. No. 87;
  Matches   33;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGC 47
              |||| ||||  |||  |||  |||  | ||||| |||| || ||||
Db        548 ATGACCAGCAAGTTGTCCGACTTGAAAACCATGGAATGACAGATGC 593


RESULT 2
BU147350
LOCUS       BU147350                894 bp    mRNA    linear   EST 03-SEP-2002
DEFINITION  AGENCOURT_8070809 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6089930
            5', mRNA sequence.
ACCESSION   BU147350
VERSION     BU147350.1  GI:22660882
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2330  row: i  column: 03
            High quality sequence stop: 713.
FEATURES             Location/Qualifiers
     source          1. .894
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6089930"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_112"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             52.1%;  Score 25;  DB 13;  Length 894;
  Best Local Similarity   75.6%;  Pred. No. 1e+02;
  Matches   31;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCA 42
              |||| ||||   ||||| |||||||||||| |  ||  ||||

Db        841 ATGAGCAGCCCCTTACCAGTGTTGTTACCAACCAAGGGGCA 881


RESULT 3
BB435111
LOCUS       BB435111                273 bp    mRNA    linear   EST 01-AUG-2000
DEFINITION  BB435111 RIKEN full-length enriched, adult pancreas Islet cells Mus
            musculus cDNA clone C820011N02 3', mRNA sequence.
ACCESSION   BB435111
VERSION     BB435111.1  GI:9274838
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
            Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
            Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
            Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
            Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A.,
            Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
            Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A.,
            Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .273
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="C820011N02"
                     /tissue_type="pancreas"
                     /cell_type="Islet cells"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult pancreas
                     Islet cells"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
```

primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I.~Islet cells were provided by Hiroo Iwata, Institute for Frontier Medical Sciences, Kyoto University, Sakyo-ku, Kyoto, 606-8507 Japan, whose assistance we gratefully acknowledge."

ORIGIN

Query Match 51.7%; Score 24.8; DB 10; Length 273;
Best Local Similarity 72.7%; Pred. No. 84;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
||||||||||||||||||||||||||||||||||||||||||||
Db 147 ATGACCACCTGGTTACCCTAGTGGTTAACCTGGAAAGAGAACTT 190

RESULT 4
AL576676/c
LOCUS AL576676 1201 bp mRNA linear EST 01-JUN-2003
DEFINITION AL576676 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI080YF21 3-PRIME, mRNA sequence.
ACCESSION AL576676
VERSION AL576676.2 GI:31314957
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 16, 2001 this sequence version replaced gi:12939053.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 1179.f For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI080CC11NP1&cluster=1179.f. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DI080CC11NP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI080YF21"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 51.3%; Score 24.6; DB 9; Length 1201;
Best Local Similarity 53.3%; Pred. No. 1.6e+02;
Matches 24; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
:::|: : || |: ||||| : ||||:||||||||:: ||:

Db 1093 WMYGRGMCGCRGKATACCCYKTAATTTACMCATGAAATGRSCCAW 1049

RESULT 5
AU299906
LOCUS AU299906 446 bp mRNA linear EST 17-JUL-2003
DEFINITION AU299906 Cryptomeria japonica developing female cone Cryptomeria japonica cDNA clone CF0701 5', mRNA sequence.
ACCESSION AU299906
VERSION AU299906.1 GI:32947357
KEYWORDS EST.
SOURCE Cryptomeria japonica (Japanese cedar)
ORGANISM Cryptomeria japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.
REFERENCE 1 (bases 1 to 446)
AUTHORS Ujino-Ihara,T., Yoshimura,K. and Tsumura,Y.
TITLE Expressed sequences in developing flower tissues of Cryptomeria japonica
JOURNAL Unpublished (2003)
COMMENT Contact: Tokuko Ujino-Ihara
Bio-resources Technology Division
Forestry and Forest Products Research Institute
Matsunosato 1, Kukizaki, Ibaraki 305-8687, Japan
Tel: 81-298-73-3211(ex.444)
Fax: 81-298-73-3795
Email: udino@ffpri.affrc.go.jp;
URL:http://www.ffpri.affrc.go.jp/labs/cjgenome/database/cjdatae.htm
l.
FEATURES Location/Qualifiers
source 1. .446
/organism="Cryptomeria japonica"
/mol_type="mRNA"
/db_xref="taxon:3369"
/clone="CF0701"
/tissue_type="developing female cone"
/clone_lib="Cryptomeria japonica developing female cone"
ORIGIN

Query Match 50.4%; Score 24.2; DB 9; Length 446;
Best Local Similarity 71.1%; Pred. No. 1.7e+02;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
||||| || | || || ||||| || | ||| | || | |||||
Db 141 GATCACCTTGATAACCCTGTTGCTAACAATGCATGGGAATATGCA 185

RESULT 6
AZ056038/c
LOCUS AZ056038 539 bp DNA linear GSS 30-MAR-2000
DEFINITION RPCI-23-413M16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-413M16, genomic survey sequence.
ACCESSION AZ056038
VERSION AZ056038.1 GI:7347274
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 539)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-413M16.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

```
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 413  row: M  column: 16
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-413M16"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
ORIGIN

  Query Match             50.4%;  Score 24.2;  DB 28;  Length 539;
  Best Local Similarity   71.1%;  Pred. No. 1.8e+02;
  Matches   32;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          4 GATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              | ||||| || || |   |||||| ||||||  |||| |||| ||
Db        493 GTTCAGATGTTTGCAATTGTTGTGACCCATTCAATGTCACAGTCA 449


RESULT 7
CB200568
LOCUS       CB200568                903 bp    mRNA    linear   EST 21-MAR-2003
DEFINITION  AGENCOURT_11271837 NICHD_XGC_Tad2 Xenopus laevis cDNA clone
            IMAGE:6875681 5', mRNA sequence.
ACCESSION   CB200568
VERSION     CB200568.2  GI:29147154
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Feb 4, 2003 this sequence version replaced gi:28233219.
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Drs. Donald Brown and Liquan Cai
             cDNA Library Preparation: CLONTECH
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3111  row: h  column: 16
            High quality sequence stop: 104.
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6875681"
                     /dev_stage="metamorphosis stage 62"
                     /clone_lib="NICHD_XGC_Tad2"
                     /note="Organ: Developing Tadpole; Vector: pDNR-LIB;
                     Site_1: SfiI; Site_2: SfiI; 5' and 3' adaptors were used
                     in cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.7 kb
                     (range 0.8-3.0 kb). 15/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA)."
ORIGIN

  Query Match             50.4%;  Score 24.2;  DB 14;  Length 903;
  Best Local Similarity   71.1%;  Pred. No. 2.1e+02;
  Matches   32;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              ||||| ||||   ||| | | ||||   |||||||  |  ||||||
Db        434 ATGATGAGCTCCCTACTCTTATTGTGCTCCATGAAGTCTCACATG 478


RESULT 8
CNS01XVV/c
LOCUS       CNS01XVV                1032 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            217C04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL172228
VERSION     AL172228.1  GI:7810285
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1032)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="217C04"
```

```
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG217BB02SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             50.4%;  Score 24.2;  DB 29;  Length 1032;
  Best Local Similarity   71.1%;  Pred. No. 2.2e+02;
  Matches   32;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 ATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATG 46
              | |  |||||||||| | | |||||||| || |  | || || ||
Db        573 AAGCACAGCTGGTTATCAGAGTTGTTACACACGTGAAGGAACCTG 529


RESULT 9
AW837303
LOCUS       AW837303                 176 bp    mRNA    linear   EST 18-MAY-2000
DEFINITION  QV2-LT0038-180300-079-d01 LT0038 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW837303
VERSION     AW837303.1  GI:7931277
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV2-LT0038-180
            300-079-d01&t3=2000-03-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 26
            High quality sequence stop: 176.
FEATURES             Location/Qualifiers
     source          1. .176
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="LT0038"
                     /note="Organ: leiomios; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             50.0%;  Score 24;  DB 10;  Length 176;
  Best Local Similarity   68.8%;  Pred. No. 1.4e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  ||||||| | |||||  ||  || ||| | ||||    ||||||
Db          4 TACAATCAGCTCGGTACCCAGGTCTTTTCCCCTCAAATCATGCATGCA 51


RESULT 10
BF634308
LOCUS       BF634308                 300 bp    mRNA    linear   EST 19-DEC-2000
DEFINITION  NF082C02DT1F1017 Drought Medicago truncatula cDNA clone NF082C02DT
            5', mRNA sequence.
ACCESSION   BF634308
VERSION     BF634308.1  GI:11898466
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula drought library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7391
            Fax: 580 221 7380
            Email: gdmay@noble.org
            Insert Length: 300   Std Error: 0.00
            Plate: 082  row: C  column: 02
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF082C02DT"
                     /tissue_type="Plantlets"
                     /dev_stage="Pooled timepoints"
                     /clone_lib="Drought"
                     /note="Vector: Lambda Zap; Contains a mixture of entire
                     plantlets harvested in a series of days-post-watering
                     timepoints. "
ORIGIN

  Query Match             50.0%;  Score 24;  DB 10;  Length 300;
  Best Local Similarity   73.2%;  Pred. No. 1.7e+02;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          5 ATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACAT 45
              | |||  |||| |   || ||| |||||||||| | ||||||
Db        221 AACAGAGGGTTGCTGTTGNTGTAACCCATGAACTAGCACAT 261


RESULT 11
BJ102563
LOCUS       BJ102563                 600 bp    mRNA    linear   EST 30-MAY-2003
DEFINITION  BJ102563 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1040a11 5', mRNA sequence.
ACCESSION   BJ102563
VERSION     BJ102563.2  GI:31238878
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
```

```
  JOURNAL   Unpublished (2002)
COMMENT     On Jan 18, 2002 this sequence version replaced gi:18245233.
            Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Caenorhabditis elegans"
                     /mol_type="mRNA"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1040a11"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"
ORIGIN

  Query Match             50.0%;  Score 24;  DB 12;  Length 600;
  Best Local Similarity   68.8%;  Pred. No. 2.2e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              |||  ||| | | ||| || | || ||| |||| ||| | ||  ||||
Db        299 TATATTCACCGGATTATCCATATTATTATCCATCAAAAGTCAATTGCA 346


RESULT 12
AL589140/c
LOCUS       AL589140                610 bp    mRNA    linear   EST 02-MAR-2001
DEFINITION  AL589140 BP Chicken Brain Library Gallus gallus cDNA clone
            ROS083H01, mRNA sequence.
ACCESSION   AL589140
VERSION     AL589140.1  GI:13194174
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Murray,F.
  TITLE     BP Chicken Brain Library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Frazer Murray
            Dept. Genomics and Bioinformatics
            Roslin Institute
            Roslin, Midlothian, EH25 9PS, UK
            Tel: +44 (0)131 527 4200
            Fax: +44 (0)131 440 0434
            Email: frazer.murray@bbsrc.ac.uk
            GCGGCCGCTTTTTTTTTTTTTTTTT 3' Poly A RNA purchased from Clonetech
            (*6854-
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9031"
                     /clone="ROS083H01"
                     /tissue_type="Brain"
                     /dev_stage="Unknown"
                     /lab_host="DH10B"
                     /clone_lib="BP Chicken Brain Library"
                     /note="Vector: pSPORT1; Site_1: Not1; Site_2: Sal1; Cloned
                     unidirectionally. Primer: Oligo dT. 5' adaptor sequence:
                     5' TCGACCTCGAG 3' ; 3' adaptor sequence: 5'
                     GCGGCCGCTTTTTTTTTTTTTTTTT 3' Poly A RNA purchased from
                     Clonetech (*6854-1)"
ORIGIN

  Query Match             50.0%;  Score 24;  DB 9;  Length 610;
  Best Local Similarity   68.8%;  Pred. No. 2.2e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  || | ||||||| | |||||||  |  | |||| || ||  ||
Db        397 TACAATGATCTGGTTATCTGTGTTGTTTACAGTAAAATAGCCCACACA 350


RESULT 13
BU294276/c
LOCUS       BU294276                766 bp    mRNA    linear   EST 27-NOV-2002
DEFINITION  603606963F1 CSEQCHN55 Gallus gallus cDNA clone ChEST590n16 5', mRNA
            sequence.
ACCESSION   BU294276
VERSION     BU294276.1  GI:25743899
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 766)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .766
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 15I"
                     /db_xref="taxon:9031"
                     /clone="ChEST590n16"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN55"
                     /note="Organ: kidney + adrenal; Vector: pBluescript II
                     KS(+); Site_1: EcoRI; Site_2: Not1; This normalized
                     library was constructed from 1 million independent clones.
                     cDNA synthesis was initiated using an oligo(dT) primer,
                     using methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

  Query Match             50.0%;  Score 24;  DB 13;  Length 766;
  Best Local Similarity   68.8%;  Pred. No. 2.4e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  || | ||||||| | |||||||  |  | |||| || ||  ||
```

```
Db        183 TACAATGATCTGGTTATCTGTGTTGTTTACAGTAAAATAGCCCACACA 136


RESULT 14
BU252275/c
LOCUS       BU252275                780 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  603748238F1 CSEQCHN38 Gallus gallus cDNA clone ChEST656110 5', mRNA
            sequence.
ACCESSION   BU252275
VERSION     BU252275.1  GI:25505294
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 780)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .780
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST656110"
                     /dev_stage="36"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN38"
                     /note="Organ: limbs; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

  Query Match            50.0%;  Score 24;  DB 13;  Length 780;
  Best Local Similarity  68.8%;  Pred. No. 2.4e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  || | ||||||| | ||||||||  |  | |||| || ||  ||
Db        331 TACAATGATCTGGTTATCTGTGTTGTTTACAGTAAAATAGCCCACACA 284


RESULT 15
BU324056/c
LOCUS       BU324056                791 bp    mRNA    linear   EST 28-NOV-2002
DEFINITION  603493436F1 CSEQCHN63 Gallus gallus cDNA clone ChEST399g10 5', mRNA
            sequence.
ACCESSION   BU324056
VERSION     BU324056.1  GI:25832057
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST399g10"
                     /dev_stage="36"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN63"
                     /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

  Query Match            50.0%;  Score 24;  DB 13;  Length 791;
  Best Local Similarity  68.8%;  Pred. No. 2.4e+02;
  Matches   33;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 TATGATCAGCTGGTTACCCGTGTTGTTACCCATGAAATGGCACATGCA 48
              ||  || | ||||||| | ||||||||  |  | |||| || ||  ||
Db        369 TACAATGATCTGGTTATCTGTGTTGTTTACAGTAAAATAGCCCACACA 322


Search completed: March 18, 2004, 20:11:24
Job time : 218.058 secs
```

THIS PAGE BLANK (USPTO)